UNIVERSITE DE NICE

UNE ETUDE COMPAREE DES MODELES D'EXPOSITION AUX SUPPORTS PUBLICITAIRES

M. JEAN-LOUIS CHANDON

1975

grenoble ort

UNIVERSITE DE NICE

UNE ETUDE COMPAREE DES MODELES D'EXPOSITION AUX SUPPORTS PUBLICITAIRES

M. JEAN-LOUIS CHANDON

Thèse pour le doctorat en sciences de gestion
présentée et soutenue en novembre 1975

SERVICE DE REPRODUCTION DES THESES DE
L'UNIVERSITÉ DES SCIENCES SOCIALES DE GRENOBLE

UNIV.
LEIDEN
BIBL.

AVERTISSEMENT

Le présent document est une reproduction faite en réduction (33 %) d'un document original fourni par l'auteur

Le Service de Reproduction des Thèses de l'Université des Sciences Sociales (Grenoble II) ne peut être tenu pour responsable de la mauvaise qualité de certaines parties. Seul l'auteur qui n'a pas eu le loisir ou la possibilité de fournir des originaux correspondant aux normes fixées par l'Atelier peut être incriminé

L'Université n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

UNE ETUDE COMPAREE DES MODELES

D'EXPOSITION AUX SUPPORTS PUBLICITAIRES

-:-:-:-:-:-

SOMMAIRE

INTRODUCTION

Présentation de la Recherche par Philip Kotler 7

CHAPITRE 1 - Présentation du Problème

1. Les deux approches du phénomène publicitaire 11
2. L'Importance de l'exposition aux Média dans la programmation commerciale 13
3. Historique de la mesure des audiences 15
4. Définition du domaine d'étude 16
5. Plan d'étude 17

CHAPITRE 2 - Revue de la littérature dans le domaine de l'exposition aux Media

1. L'état de l'art en matière de plan publicitaire
   11. Définition de la planification publicitaire . . . 23
   12. Les complexités de la planification publicitaire . 24
   13. Tout d'horizon des principaux modèles de planification publicitaire 45
2. Les Modèles d'exposition aux Media
   21. Le modèle implicite 49
   22. Le modèle naïf 50
   23. Le modèle subjectif 52
   24. Les modèles sophistiqués 54

CHAPITRE 3 - Les Objectifs de l'Etude

1. Le Regroupement et la description de tous les modèles connus d'exposition aux supports publicitaires . . . 65
2. La Recherche et la classification des hypothèses des modèles 66

3. La construction d'une grille d'évaluation à critères multiples 66
4. Le test et l'évaluation des modèles à l'aide de banques de données unifiées 67
5. Le développement de nouveaux modèles d'exposition aux supports publicitaires 68
6. L'évaluation et le classement des modèles d'exposition aux supports publicitaires 68

Première Partie : LES CONCEPTS DE BASE

CHAPITRE 4 - Définition et Mesures

1. La Circulation 72
2. La pénétration 74
3. L'exposition 75
   31. Les média imprimés
      311. Le choix de la technique de collecte des données 77
      312. Problèmes liés au choix du critère de lecture 78
      313. Problèmes liés à la formulation des questions 80
      314. Les méthodes utilisées en France et aux Etats-Unis 90
      315. Différences attribuables à la technique de collecte des données 92
      316. Différences attribuables à la formulation des questions et aux critères de lecture 93
   32. Les Media radiodiffusés 97
4. L'audience 111
5. La Couverture nette 113
6. L'accumulation des audiences 117
   61. Les mesures directes 117
   62. Les mesures indirectes 118
7. La duplication des audiences 119
8. La matrice d'exposition 121
9. La distribution des occasions de voir 123
10. Audiences fidèles et audiences infidèles 125
11. Conclusions 128

Deuxième Partie : LES MODELES D'EXPOSITION

CHAPITRE 5 - Les Modèles d'Accumulation d'Audience

1. La relation fondamentale liant la couverture nette cumulée à la distribution des occasions de voir 133
   12. De la distribution des occasions de voir à la couverture nette cumulée 135
   13. De la couverture nette cumulée à la distribution des occasions de voir 138

2. Les relations liant la fidélité au support à l'accumulation et à la distribution des occasions de voir . . . 143
   21. De la couverture nette cumulée à la distribution des audiences fidèles 144
   22. De la distribution des audiences fidèles à la couverture nette cumulée 145
   23. De la distribution des occasions de voir à la distribution des audiences fidèles 147
   24. De la distribution des audiences fidèles à la distribution des occasions de voir 148

3. Les modèles d'accumulation d'audience
   31. Les modèles Ad-Hoc 153
      311. L'accumulation logarithmique 154
      312. L'accumulation exponentielle 158
      313. La formule d'Agostini 160
      314. La courbe d'accumulation d'Engelman 165
      315. La formule d'Hofmans 166
   32. Les modèles stochastiques 173

MODELES D'ORDRE ZERO

      321. Le modèle à population homogène : le modèle binomial 175
      322. Les modèles d'accumulation à population segmentée 181
      323. Les modèles d'accumulation à hétérogénéité individuelle 191
         - Le modèle Politz 196
         - Le modèle Bêta Binomial 200
         - Le modèle négatif Binomial . . . 229

324. Les modèles d'accumulation à double hétérogénéité 241
- Le modèle de Broadbent 241
- Le modèle de Morgensztern 247
- Le Modèle Bêta Binomial segmenté 261

MODELE D'ORDRE UN

325. Le modèle d'accumulation Binomial Markovien . . 270
33. Les modèles de simulation 279
331. L'affectation actuarielle des probabilités . . 281
332. L'affectation fixe des probabilités 286
333. L'affectation empirique des probabilités . . . 287
334. L'affectation personnelle des probabilités . . 290
335. L'affectation au dernier numéro 300

4. Evaluation critique des modèles d'accumulation d'audience 314
41. Les quinze critères d'évaluation 315
42. Le classement des 21 modèles d'accumulation d'audience 316
5. Conclusions 330

CHAPITRE 6 - Les Modèles de Duplication d'Audiences

1. Les modèles Ad-Hoc
11. La formule d'Agostini 339
12. La formule d'Hofmans 347
13. La formule de Starch 349
14. Le modèle de Young 354
15. La formule de Little et Lodish 355

2. Les modèles stochastiques 359
21. Le Modèle Binomial 361
22. Le Modèle de Sainsbury 365
23. Le Modèle de Kwerel 369
24. Le Modèle de Sainsbury Modifié 375
25. Le Modèle Goodhart-Ehrenberg 379
26. Les Modèles d'Agrégation Séquentielle 380
27. Le Modèle d'indépendance conditionnelle 385

28. Le Modèle d'Hofmans étendu 388
29. Le Modèle de Metheringham 391
30. Le Modèle de Benson 402

3. Evaluation et classement des modèles de duplication d'audience 406
4. Conclusions 409

CHAPITRE 7 - Les Modèles complets à insertions multiples dans plusieurs Véhicules

1. Généralisation à plusieurs véhicules de la relation entre couverture nette et distribution des occasions de voir 413
11. Une relation récursive pour les matrices d'exposition 414
12. Des matrices d'exposition aux matrices de couverture nette 416
13. Des matrices de couverture nette aux matrices d'exposition 420
2. Les modèles stochastiques 424
21. Modèles à population homogène 425
211. duplication Aléatoire 425
. Le modèle Binomial 425
. Le modèle Binomial Multivarié avec indépendance des véhicules 430
212 duplication non Aléatoire 435
. Le modèle Multinomial 435
. Le modèle Binomial Markovien 443

22. Modèles à population hétérogène
221. Modèles Distributionnels
. Le modèle de Metheringham 458
. Le modèle Bêta Matriciel 467
. Le modèle Multinomial composé . . . 478
222. Modèles d'Approximation
. Le modèle Modal. 491
. Le modèle d'Hofmans 495
. Le modèle de Morgensztern 500
. Le modèle de casualité constante . . 507
. Le modèle Metheringham Multinomial Composite 520

3. Les Modèles de Simulation 522
31. Le modèle des probabilités personnelles 523
32. Le modèle C.A.M. 526
33. Le modèle S.C.A.L. 529
4. Evaluation Critique des modèles complets 532
41. Les critères d'évaluation 534
42. Le classement des modèles complets 536
5. Conclusions et directions pour les recherches futures 537

ANNEXES - A 544
B 546
C 547
D 551
E 552
F 581
Fin des Annexes 599

PRESENTATION DE LA RECHERCHE

par

Philip KOTLER

Le plus grand nombre d'applications des ordinateurs aux problèmes du marketing se situe probablement dans le domaine de la planification des supports publicitaires. Cependant, les complexités du problème et le manque de connaissances de base concernant la réponse du consommateur aux messages publicitaires, ont empêché jusqu'à ce jour, et sont encore susceptibles d'empêcher dans un futur proche, l'avènement d'une solution mécanique.

La prévision du nombre d'individus exposés aux supports porteurs du message de la campagne constitue le point de départ nécessaire de tout modèle de planification des media. L'étude de ce problème est le sujet de cette recherche qui réalise la première comparaison exhaustive des performances de 47 modèles d'exposition aux supports publicitaires.

L'objectif est de déterminer quels sont les meilleurs modèles d'exposition aux supports publicitaires actuellement disponibles. Cet objectif a été atteint par la recherche et la classification des hypothèses de chaque modèle, par l'élaboration d'une grille d'évaluation à critères multiples, et par la mise à l'épreuve des modèles à l'aide de banques de données homogènes.

Au cours de cette opération, onze nouveaux modèles d'exposition ont été développés, et les performances des modèles antérieurs ont été améliorés, tant dans le domaine de l'accumulation que dans celui de la duplication des audiences.

Cette recherche débute par une revue de la littérature des vingt dernières années qui est utilisée pour présenter l'état des connaissances dans le domaine de la planification des média, et pour établir une distinction entre les modèles d'exposition implicite, naif, subjectif, et sophistiqué. Puisque la signification des concepts publicitaires essentiels varie d'un pays à un

autre et d'un support à un autre, l'auteur présente sa propre gamme de définitions, accompagnés d'une description des mesures grâce auxquelles les concepts sont rendus opérationnels. Le sujet fort controversé de la mesure des audiences est discuté. Les mérites relatifs des méthodologies courantes sont évalués. Les divergences entre les estimations produites par les diverses sociétés de recherche sont illustrées pour les magazines, les journaux, les stations de radio, et les programmes de télévision.

L'étude de 47 modèles d'exposition qui constitue le coeur de cette recherche est divisée en trois parties :

- dans la première, 21 modèles d'accumulation d'audience sont examinés. Un traitement explicite des relations mathématiques liant l'accumulation d'audience à la fréquence d'exposition et à la fidélité aux supports est fourni. Le problème de la stabilité des probabilités d'exposition est étudié grâce à la présentation d'un modèle markovien et d'une version modifiée du modèle de Schreiber à probabilités variables. Ensuite, le pour et le contre des cinq méthodes d'affectation des probabilités utilisées pour la simulation de l'accumulation d'audience sont évaluées selon les critères de reproducibilité des données, d'unicité de la solution et de robustesse de la méthode.

- dans la seconde partie, six modèles de duplication ad hoc et onze modèles de duplication stochastiques sont examinés. Leurs performances sont évaluées sur la base de quinze critères.

- dans la troisième partie, la relation existant entre l'accumulation d'audience et la fréquence d'exposition est étendue au cas de plusieurs véhicules. Une équation de récurrence générale est établie qui permet de déduire la matrice d'exposition d'un plan réduit, à partir du plan d'insertion complet. Les quinze modèles complets, traitant des plans d'insertion multiples dans plusieurs véhicules, sont classifiés en modèles à population homogène qui ne tiennent pas compte des différences individuelles, en modèles à population hétérogène, qui tiennent compte des différences individuelles dans les probabilités d'exposition, et en modèles de simulation. Leurs performances sont évaluées sur la base de vingt et un critères découlant de considérations théoriques, empiriques et pratiques.

Cette recherche, entièrement originale, consitue la première étude comparée d'ensemble des performances des modèles d'exposition aux supports publicitaires. Cet effort de validation s'accompagne de la production d'une

classification systématique qui permet de comprendre pourquoi et quand certains modèles fonctionnent mieux que d'autres. La troisième contribution de l'auteur réside dans la construction d'un nouveau modèle d'accumulation, le modèle Béta Binomial Segmenté, et d'un nouveau modèle complet, le modèle Multinomial Composite, qui possède des performances supérieures à celles du modèle de Metheringham tout en ne nécessitant pas de données supplémentaires. La quatrième contribution tient à la création d'une grille d'évaluation à critères multiples qui se révèlera utile pour le classement et la sélection des modèles d'exposition aux supports publicitaires, existant et à venir.

La question de l'exposition au message publicitaire n'est pas traitée dans cette recherche. Elle sera vraisemblablement l'occasion de progrès importants dans les dix prochaines années. Nous espérons que la présente contribution à la validation des modèles d'exposition aux supports publicitaires ouvrira la voie à la construction de ces modèles d'avenir.

INTRODUCTION

CHAPITRE I

PRESENTATION DU PROBLEME

1. Les deux approches du phénomène publicitaire

Les dépenses publicitaires françaises en 1974 s'élevaient à plus de 7 milliards de francs selon l'estimation de l'IREP (1). Aux Etats-Unis, la même année, 26,5 milliards de dollars étaient investis dans les huit principaux media selon l'estimation de R.J.COEN (2). Pour l'ensemble des pays développés on peut estimer que le total des sommes dépensées avoisine 45 millions de dollars.

La table n°1 décrit l'évolution des dépenses publicitaires, de 1964 à 1970, dans les principaux pays du monde occidental. Les Etat-Unis arrivent largement en tête, tant en valeur absolue qu'en dépenses par tête d'habitant. On notera toutefois que la Suisse dépense davantage en pourcentage de son produit national brut. L'Allemagne de l'Ouest se place en seconde position et enregistre, avec le Japon, l'accroissement le plus rapide des dépenses publicitaires. En dehors de ces trois pays, les dépenses sont plus modestes, aussi bien en Angleterre, qu'en France ou au Canada.

La France, avec l'Italie, s'attribue la plus faible dépense en pourcentage du produit national brut. Néanmoins, les dépenses publicitaires y atteignent trois fois le budget du Ministère de la Justice.

TABLE 1

Evolution des Dépenses Publicitaires
dans les pays ayant les plus fortes dépenses

PAYS	1964			1970		
	Millions de $	% PNB	$ par tête	Millions de $	% PNB	$ par tête
ETATS-UNIS	14 155	2,2	73,7	19 600	2,1	95,7
ALLEMAGNE	1 278	1,2	21,7	2 694	1,6	43,7
JAPON (1966)	955	1,3	9,7	2 115	1,1	20,4
ANGLETERRE	1 134	1,2	20,9	1 265	1,1	22,7
CANADA (1966)	738	2,0	36,9	1 037	1,2	48,5
FRANCE	473	0,5	9,8	996	0,7	19,6
ITALIE	295	0,6	5,7	489	0,6	9,1
PAYS-BAS	214	1,3	17,6	410	1,4	31,3
SUISSE	167	1,3	28,4	428	2,2	68,2

Source : Pierre Kendé 1969 and International Advertising Association 1972

L'ampleur de ces sommes soulève deux types de questions. On peut s'interroger sur les conséquences économiques et sociales qui résultent des dépenses publicitaires. C'est le point de vue de l'économiste. On peut aussi se soucier de déterminer l'allocation, dans le temps et dans l'espace, du budget publicitaire qui contribuera le mieux au profit de la firme. C'est le point de vue du gestionnaire. Notre recherche adopte résolument le second point de vue.

Les deux approches ne sont certes ni indépendantes ni incompatibles, cependant, la recherche commerciale a été délaissée pendant fort longtemps par les chercheurs universitaires. Le dédain, voire l'hostilité, manifesté à son égard a créé une indéniable coupure entre les pratiques mises au point sur le tas et la théorie de la firme.

Nous nous efforçons ici de contribuer à la disparition de cette coupure en étudiant un aspect important du phénomène publicitaire : l'exposition aux media. Notre méthodologie vise à la construction de modèles théoriques rigoureux du phénomène, ainsi qu'à leur validation expérimentale. Nous espérons ainsi contribuer à une meilleure connaissance du phénomène tout en mettant à la disposition des praticiens des moyens opérationnels de prise de décision.

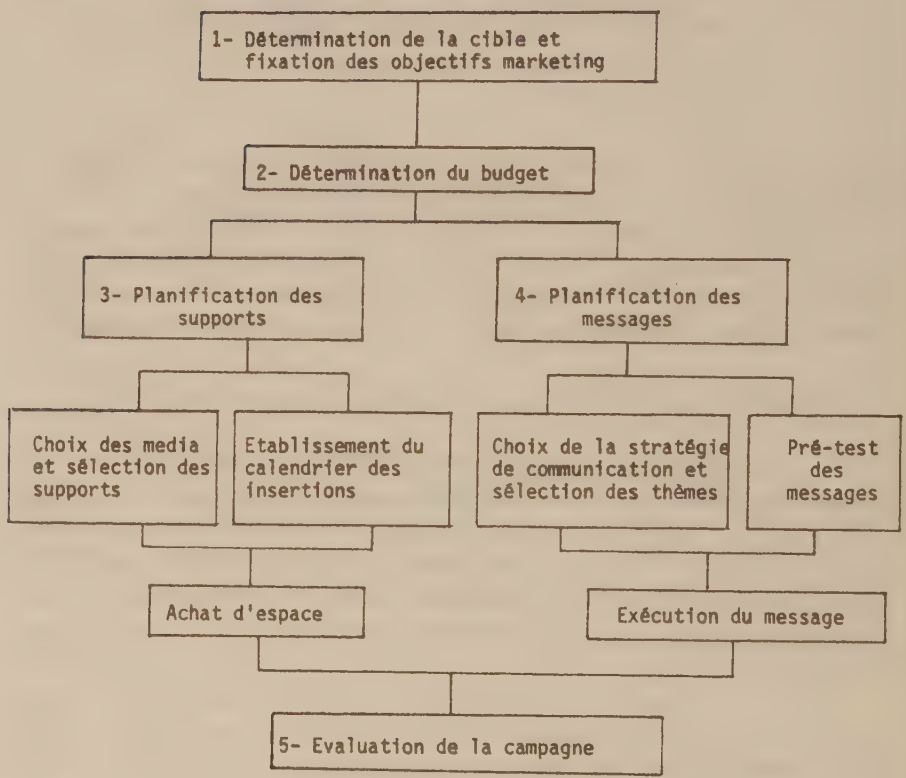
2. L'Importance de l'Exposition aux Media dans la Programmation Commerciale

La planification des media constitue un élément essentiel de la programmation commerciale. Elle consiste à choisir la combinaison de supports publicitaires, l'assortiment des messages et le calendrier de leurs insertion qui contribueront au mieux à la réalisation des objectifs de vente de la firme. Elle s'insère donc parmi les autres éléments de la "pression marketing"(3) de la firme qui sont le produit lui-même, son prix et sa position dans les réseaux de distribution.

Nous n'examinerons pas ici comment fixer la pression marketing optimale en fonction de la structure du marché, des caractéristiques de la demande et de la stratégie de la firme. D'excellentes études ont déjà été réalisées en ce domaine, notamment par Dorfman et Steiner (4), P. Kotler (5), J.J. Lambin (6) et C. Palda (7). Nous nous préoccuperons exclusivement de la question suivante: comment atteindre les bons consommateurs, au bon moment et le bon nombre de fois avec un minimum de gaspillage. Cette tâche, qui confronte les spécialistes de la planification publicitaire, peut être décomposée en cinq étapes qui sont présentées dans la table n°2.

TABLE 2

LES CINQ ETAPES DE LA PLANIFICATION PUBLICITAIRE



Notre recherche est une contribution à la planification des campagnes publicitaires, mais elle n'en recouvre pas tous les aspects. Nous ne traiterons ni de la fixation des objectifs et des budgets, ni le la planification des messages. Notre étude ne recouvrira même pas la totalité des problèmes liés à la sélection des média (8) et à la chronologie des insertions. Nous traiterons uniquement de cette partie vitale de tout modèle de planification des média qui a pour objet la prévision du nombre d'individus exposés aux supports publicitaires porteurs des messages de la campagne, ainsi que la prévision de la fréquence d'exposition à ces supports.

Cette question préliminaire est véritablement d'une importance primordiale, puisque toutes les décisions concernant la campagne reposent sur une évaluation correcte de l'importance et des caractéristiques de l'audience atteinte par une combinaison donnée de supports. En l'absence d'informations sur l'exposition aux supports, il est impossible de repérer les consommateurs-cibles, de créer un thème approprié, d'évaluer l'exposition aux messages, de décider de la combinaison de supports la plus appropriée, ainsi que la chronologie des insertions ou de la taille du budget à dépenser. Si l'on ne connaît pas la fréquence d'exposition produite par un plan de campagne, il est impossible d'évaluer les effets de la répétition sur la connaissance du produit par le consommateur, sur ses attitudes et sur ses achats, et, en conséquence, aucun contrôle de l'efficacité de la campagne publicitaire n'est possible.

3. Historique de la Mesure des Audiences

La nécessité de posséder des données sur les supports ainsi que des modèles pour les analyser a été reconnue depuis longtemps. Le premier service de recherche de marketing au Etat-Unis fut créé il y a 60 ans par Charles C. Parlin du Saturday Evening Post (9). La première mesure systématique de la circulation des supports imprimés commença en 1914, aux Etat-Unis, avec la création de l'Audit Bureau Of Circulation (10), suivie peu après, en France, par la création de l'Office de Justification de la Diffusion.

Cependant, il fallut attendre jusqu'en 1938 pour voir apparaître la première recherche sur les supports basée sur le concept d'exposition (11). Elle fut réalisée par Crossley pour le magazine Life (12) et fut suivie, un an plus tard, par la première étude de l'IPA sur l'audience de la presse en Angleterre (13). Ces deux études ouvrirent la voie aux recherches portant sur l'exposition aux supports en insistant non plus sur le nombre d'unités physiques

distribuées, mais sur le nombre d'individus entrant en contact avec le support publicitaire.

Les études sur l'exposition aux supports publicitaires prolifèrèrent rapidement pendant la période d'après guerre. L'introduction de la télévision et la concurrence qui s'ensuivit avec les média établis alimenta une course folle à qui aurait la plus grande audience entre les chaînes de télévision et les principaux magazines. Une des conséquences de cette surenchère fut la multiplication des données quantitatives,qui conjuguée avec le succès de techniques de recherche opérationnelle dans d'autres domaines de gestion, entraîna le développement de modèles quantitatifs dans le domaine de la sélection des supports (14). Cependant les spécialistes de la recherche opérationnelle n'avaient pas, à cette époque, une claire compréhension des complexités du problème de la planification publicitaire. De même, les spécialistes de la planification publicitaire ne disposaient pas des bases mathématiques nécessaires à la compréhension des limitations de la programmation linéaire. Cette première tentative était vouée à l'échec. Elle eut malgré tout un résultat bénéfique incontestable : la recherche de modèles viables d'exposition aux supports publicitaires.

Quinze ans plus tard, et après la création de dizaines de modèles, le besoin d'une structure unifiée pour comparer, évaluer et valider les modèles de planification des média, se fait clairement sentir au sein de la communauté des publicitaires (15).

4. Définition du Domaine d'Etude

Cette dissertation tente, modestement, de fournir des réponses dans le domaine des modèles d'exposition aux supports publicitaires. Le problème qui sera abordé peut être énoncé très simplement : quels sont les meilleurs modèles d'exposition aux supports publicitaires actuellement disponibles ? Cette question suggère immédiatement deux autres questions : qu'est-ce qu'un modèle d'exposition aux supports, et quels sont les critères selon lesquels il peut être évalué ? On peut définir un modèle d'exposition aux supports en se référant à sa fonction ou en se référant à sa structure. La fonction d'un modèle d'exposition est d'estimer combien d'individus sont exposés combien de fois aux supports dans lesquels un message publicitaire a été inséré un nombre donné de fois. Cette estimation est réalisée grâce à une combinaison particulière de données et d'hypothèses qui définissement la structure du modèle. Cette

structure n'a pas besoin d'être définie mathématiquement, pas plus que les hypothèses n'ont besoin d'être énoncées explicitement. Un modèle subjectif reste toujours un modèle. Cependant il n'est pas forcément bon, ce qui nous amène à poser notre seconde question : quels sont les critères d'évaluation ? Fondamentalement, ils sont de trois sortes : théorique, empirique et pratique. Les critères théoriques permettent de dire si oui ou non les exigences de cohérence interne, d'économie et de falsifiabilité des hypothèses sont satisfaites. Les critères empiriques permettent de dire si oui ou non le modèle remplit sa fonction, en évaluant sa robustesse, la qualité de ses prédictions et l'étendue de son domaine d'application. Les critères pratiques permettent de dire si le modèle est efficace du point de vue de l'analyse numérique et si les données qu'il requiert peuvent être obtenues à un prix compatible avec la valeur informative des résultats du modèle.

Pourquoi se concentrer sur le problème de l'exposition aux supports ? Les modèles d'exposition aux supports ne répondent pas à la question qui est probablement de la plus haute importance pour le spécialiste de la publicité, à savoir combien d'individus sont exposés au message publicitaire. Cependant, afin de donner une réponse satisfaisante à cette question très importante, il faut tout d'abord résoudre le problème de l'exposition aux supports. La raison en est qu'une exposition au message ne peut se produire que si une exposition au support se produit également. Néanmoins, le contraire n'est pas vrai, puisqu'on peut très bien être exposé à un support publicitaire sans être en même temps exposé aux messages publicitaires qu'il contient. Autrement dit, l'exposition au support est une condition nécessaire mais non suffisante de l'exposition au message. Par conséquent, il faut tout d'abord examiner le problème de l'exposition aux supports.

Malgré l'importance stratégique du problème de l'exposition aux supports, il n'existe aucune recherche comparant l'ensemble des modèles d'exposition. Notre recherche est donc entièrement originale, puisqu'elle réalise la première étude comparative d'ensemble des performances de 47 modèles d'exposition aux supports publicitaires.

5. Plan d'Etude

Notre recherche débute par une revue de la littérature des 20 dernières années. Cette revue couvre un total de 116 références dans les domaines de l'exposition aux supports et de la planification des média. Nous

puisons abondamment à ces sources dans le chapitre 2 pour décrire l'état de l'art en matière de planification des média, et pour mettre en évidence les caractéristiques que devrait posséder un modèle idéal. Cette classification sert d'introduction aux concepts fondamentaux d'accumulation d'audience et de duplication d'audience. Ayant ainsi, d'une manière critique, passé en revue les approches de base employées pour construire les modèles d'exposition aux supports, et relevé les principales tendances, nous définissons ensuite dans le chapitre 3 les objectifs de notre recherche et présentons notre plan de recherche ainsi que nos banques de données.

Le chapitre 4 analyse les principaux concepts liés à la mesure des audiences et discute les différents problèmes de mesure liés à la collecte des données d'audience Dans le domaine des média imprimés les mérites relatifs des principales méthodologies couremment employées sont inventoriés et évalués sur la base de critères théoriques et empiriques. Une étude similaire est menée concernant les média radiodiffusés, et des données sont fournies pour illustrer les limites des méthodes de mesure. Puis nous présentons une définition rigoureuse des concepts d'accumulation et de duplication, suivie de l'introduction de deux nouveaux concepts, de fidélité et d'instabilité des audiences.

Les trois chapitres suivants forment la partie principale de la dissertation. Dans le chapitre 5 nous analysons la relation qui, dans le cas d'un seul support, existe entre l'accumulation de l'audience à travers le temps, la fidélité des audiences, et la distribution des occasions de voir. Puis nous analysons et décrivons 21 modèles d'accumulation dont l'objectif est de prédire le nombre d'individus exposés à un nombre donné de parutions d'un seul support. Cinq de ces modèles sont entièrement nouveaux, et leurs performances sont comparées avec celles des 16 modèles existants, et ceci sur la base de 15 critères d'évaluation.

Dans le chapitre 6 nous décrivons et analysons 14 modèles de duplication visant à prédire le nombre d'individus exposés à une combinaison de supports publicitaires, lorsqu'il n'y a pas plus d'une insertion par support. Quatre de ces modèles ont été développés par l'auteur de cette thèse et leurs performances sont évaluées en même temps que celles des 10 modèles préexistants.

Dans le chapitre 7, nous élargissons la relation entre l'accumulation d'audience et la distribution des occasions de voir au cas de supports multiples. Nous analysons ensuite 15 modèles à insertion multiples dans plusieurs supports

dont 4 sont entièrement nouveaux, et nous comparons leurs performances sur la base de 18 critères d'évaluation.

Enfin, nous résumons les principales conclusions de notre étude et discutons de voies nouvelles pour la recherche.

NOTES

(1) - Institut de Recherches et d'Etudes Publicitaires, Paris, 1975

(2) - COEN F.6. "1975 Ad volume may top $ 28 billions" Advertising Age, 16 Décembre 1974

(3) - Traduction de l'expression "Marketing Mix"

(4) - DORFMAN R. et STEINER P.O. "Optimal advertising and optimal quality", The American Economic Review, vol 44, 1954, pp. 826-836

(5) - KOTLER P. "Competitives strategies for new product marketing over the life cycle", Management Science, Décembre 1965, pp.104-119

(6) - LAMBIN J.J. Modèles et Programmes de Marketing, Presses Universitaires de France, ("Systêmes-Décisions") , 1970

(7) - PALDA K.S. Economic Analysis for Marketing Decision, Prentice Hall, 1969

(8) - Par média on entend un groupe de supports publicitaires homogènes tels que les magazines ou les journaux ou les stations de radio et de télévision, ou encore la publicité extérieure.

(9) - RAMOND C.K. The Art of Using Science in Marketing, Harper and Row, New-York, 1974

(10) - Bureau d'Etude de la Circulation, 123 North Wacker Drive, Chocago

(11) - Le Concept d'exposition est défini avec précision dans le chapitre 4

(12) - Crossley Surveys, A continuing study of Magazine Audiences, Report N°1, LIFE, New-York, Décembre 1938

(13) - Institute of Practitioner in Advertising, Survey of Press Readers hip, London, 1939

(14) - Anonyme, "Young & Rubicam, BBDO, unleash media computerization" Advertising Age, October 1962

(15) - HALEY R.I. "Do we really know what we are doing ?" Proceedings of the American Marketing Association Conference 1971, pp. 217-219

CHAPITRE 2

REVUE DE LA LITTERATURE DANS LE DOMAINE DE L'EXPOSITION AUX MEDIA

Puisque les modèles d'exposition aux supports publicitaires ne sont qu'un élément des modèles de planification publicitaire, il est utile de commencer par une évaluation de l'état de l'art en matière de planification publicitaire. Nous présenterons ensuite une évaluation des modèles d'exposition stricto sensu.

1. L'Etat de l'Art en matière de plan publicitaire.

Dans ce paragraphe, nous présentons une définition de la planification publicitaire et nous énonçons les caractéristiques que devrait posséder un modèle idéal de planification publicitaire. Ce schéma de modèle idéal est utilisé ensuite pour classer et évaluer les modèles de planification les plus remarquables de ces quinze dernières années.

11. Définition de la planification publicitaire.

Un modèle de planification publicitaire est un ensemble de procédures, programmées ou non, ayant pour objet le choix de la combinaison de supports publicitaires, de la nature et du nombre de messages à placer dans chacun d'eux, ainsi que du calendrier des insertions, qui maximisent une fonction objective donnée, dans les limites imposées par le budget et la disponibilité des supports.

Sur la base de cette définition, les modèles de planification publicitaire peuvent être classés en trois grandes familles :

a) - Les modèles de sélection des supports.

Ils permettent de choisir une combinaison de supports appartenant à un ou plusieurs media. Leur produit final est une liste de supports.

b) - Les modèles de calendrier d'insertion.

Ils permettent d'établir un calendrier d'insertion, pour un ou plusieurs supports et pour la durée de la campagne. Leur produit final est une liste, support par support, des dates d'insertion de chacun des messsages.

c) - Les modèles de planification publicitaire.

Ces modèles s'efforcent de résoudre simultanément le problème de la sélection des supports et celui de la chronologie des insertions.

Idéalement, un modèle de planification publicitaire réellement complet devrait aussi permettre de choisir entre divers messages publicitaires. Toutefois, l'étude des interactions messages-media-chronologies constitue un terrain encore relativement peu exploré quoique très prometteur. La traditionnelle séparation, existant dans les agences de publicité entre le département media-planning et le département créativité, constitue un obstacle majeur dans ce domaine.

Le choix d'une stratégie de communication et sa concrétisation en un portefeuille de messages n'est certes pas indépendant du choix des supports. Toutefois, en pratique, la liaison message-support se manifeste surtout sous formes de contraintes imposées par le département créatif et visant à l'exclusion de tel ou tel support.

12. Les complexités de la planification publicitaire.

Le problème de la sélection des supports et de l'établissement du calendrier des insertions est extrêmement complexe car il doit tenir compte de huit types de contraintes qui réagissent les unes sur les autres. Ces contraintes sont présentées dans la table 3 qui décrit les huit décisions principales qui doivent nécessairement être prises pour arriver à un plan de campagne efficace et cohérent.

TABLE 3

LES HUIT COMPLEXITES DE LA PLANIFICATION DES SUPPORTS

DECISIONS A PRENDRE	DESCRIPTION DU PROBLEME
PONDERATION DES OBJECTIFS	Déterminer le dosage optimal entre couverture et répétition ; continuité et alternance ; sélectivité et ubiquité ; concentration et dilution.
PONDERATION DES AUDIENCES	Définir et pondérer les segments du marché selon leur taille et leur potentiel, puis assortir les supports aux segments.
PONDERATION DES EXPOSITIONS	Estimer les probabilités d'exposition aux supports et les probabilités d'exposition aux messages, support par support et période par période, compte tenu de l'accumulation, de la duplication et de la saisonnalité des audiences.
PONDERATION DES SUPPORTS	Assortir les supports aux messages compte tenu de la nature du produit, du climat éditorial et des caractéristiques propres du support.
PONDERATION DES MESSAGES	Estimer l'influence du format (taille, couleur, structure, ...), du contenu (thème, exécution, ...) et de la position (couverture, voisinage, ...) de l'annonce sur la perception, la crédibilité, la rétention et l'efficacité du message.
PONDERATION DES FREQUENCES	Intégrer, sous forme d'une courbe de réponse, l'influence des campagnes antérieures, les rendements décroissants et la saturation résultant d'expositions répétées avec l'oubli qui prend place en l'absence d'exposition, sans pour autant négliger la saisonnalité des ventes et la fréquence d'achat du produit.
PONDERATION GEOGRAPHIQUE	Déterminer la répartition optimale des supports et des messages par régions, compte tenu de leurs tailles, de leurs potentiels et de leurs particularités socio-culturelles.
PONDERATION DES COMPETITEURS	Prévoir les stratégies des concurrents et se réserver une marge de manoeuvre pour des actions ponctuelles.

L'examen de cette table fait apparaître l'interdépendance des décisions à prendre et souligne la nécessité de construire des modèles capables de prendre en compte simultanément les aspects probabilistes non linéaires, interactifs et dynamiques de la planification des media.

Le premier problème qui se pose est celui de la pondération des objectifs assignés à la campagne. Il faut tout d'abord décider de l'importance relative à accorder à la couverture nette vis-à-vis de la fréquence moyenne d'exposition. Les principes suivants peuvent guider ce choix délicat :

- Un haut degré de couverture est nécessaire pour le lancement d'un produit nouveau ou la mise sur pied d'une promotion spéciale.

- Une haute fréquence d'exposition est nécessaire lorsque le message est complexe, lorsque le produit est fréquemment acheté et lorsque la fidélité à la marque est faible.

- Pour un budget donné, on ne peut pas accroître simultanément la couverture et la fréquence sans diminuer le format ou la durée du message.

- Au delà d'un certain seuil de couverture, variable d'un media à un autre, les achats d'espace supplémentaire contribuent surtout à accroître la fréquence d'exposition.

- La couverture maximale s'obtient en utilisant plusieurs supports s'adressant à des publics aussi variés que possible.

- La fréquence maximale s'obtient en concentrant les achats d'espace dans un petit nombre de support sélectifs.

La table suivante, basée sur les tarifs des chaînes de télévision américaine, illustre ces principes en montrant comment un budget donné de $ 300 000 peut produire une grande variété de couvertureset de fréquences selon la durée du film publicitaire.

TABLE 4

UN EXEMPLE DE DOSAGE

DUREE - COUVERTURE - FREQUENCE

POUR UNE CAMPAGNE TELEVISEE

Budget = $ 300 000

	Plan 1	Plan 2	Plan 3	Plan 4	Plan 5
DUREE	1 message de 6 minutes	2 messages de 3 minutes	6 messages de 1 minute	12 messages de 30 secondes	60 spots de 10 secondes
COUVERTURE NETTE	22%	36 %	65 %	78 %	85 %
FREQUENCE MOYENNE	5,0	3,8	2,4	4,0	5,2
GRP	110	138	156	312	438

Le second choix stratégique est celui du dosage continuité-intermittence. Les divers facteurs susceptibles de guider ce choix ont été résumés dans la table 5.

Il apparait clairement qu'il n'y a pas de réponse toute faite à ce dilemme. Il convient de tenir compte de la nature du produit, de sa fréquence d'achat, de la saisonnalité des ventes, de la part de marché atteinte, des stratégies de la concurrence et du budget disponible.

Le troisième choix stratégique est celui qu'il faut opérer entre la concentration dans un unique media et la diversification dans plusieurs media. La table 6 recense les avantages et inconvénients de chacune des deux options.

Une fois les divers choix stratégiques effectués, les choix techniques concernant la pondération des audiences, des expositions, des supports, des messages, des fréquences, des régions géographiques et les actions de la concurrence, seront pris dans le cadre d'un modèle de planification publicitaire. Nous allons donc maintenant passer en revue les divers types de modèles disponibles.

Les modèles de planification publicitaire peuvent être analysés du point de vue de leur philosophie, de leur technique, du domaine qu'ils couvrent et des critères de choix qu'ils utilisent. Ces divers aspects sont détaillés dans la table 7 qui présente une vue d'ensemble des modèles de planification publicitaire.

M. Broadbent (1) propose de distinguer entre les modèles d'optimisation d'une part, et les modèles d'évaluation d'autre part. Les premiers cherchent à déterminer le plan publicitaire qui maximise un critère prédéterminé, tel le coût aux mille lecteurs utiles, étant donné un ensemble de contraintes à respecter. Les seconds ne produisent pas de plan publicitaire, ils se contentent d'évaluer divers plans, produits en dehors du modèle, et de les classer selon un ou plusieurs critères de performance prédéterminés. Les modèles High Assay, Mediac et Admod sont représentatifs de la première philosophie, tandis que les modèles CAM, Electre, SCAL & Ad.Me Sim exemplifient la seconde. Certains modèles tels que COSMIC et IMS font preuve d'une plus grande versatilité en permettant à l'utilisateur de passer séquentiellement de l'évaluation à l'optimisation.

L'approche optimisante, la plus ambitieuse, souffre de l'inéluctable limitation du nombre de paramètres pouvant être pris en considération par l'algorithme d'optimisation. De son côté, l'approche évaluative, quoique plus riche

TABLE 5

DETERMINATION DU DOSAGE : CONTINUITE - INTERMITTENCE

ARGUMENTS EN FAVEUR DE LA CONTINUITE

- Combattre l'oubli.
- Bénéficier d'effets cumulatifs.
- Accompagner la continuité des ventes.
- S'opposer aux efforts continus des concurrents.
- Stimuler les achats répétés.
- Accroître la fidélité à la marque.
- Renforcer les opinions du client quant à l'excellence de son choix.
- Obtenir de meilleurs ristournes de la part des supports.

ARGUMENTS EN FAVEUR DE L'INTERMITTENCE

- Atteindre le seuil de perception.
- Frapper un grand coup.
- Suivre la saisonnalité des ventes.
- Contrecarrer les efforts saisonniers des concurrents.
- Stimuler les achats impulsifs.
- Attaquer les marques concurrentes.
- Soutenir une promotion spéciale.
- Atteindre le seuil minimum de dépense permettant l'accès à la télévision.

ARGUMENTS CONTRE LA CONTINUITE

- Elle engendre la saturation, la fatigue et l'irritation des individus.
- Elle nécessite un budget important.

ARGUMENTS CONTRE L' INTERMITTENCE

- Elle affaiblit la position concurrentielle pendant les périodes de silence.
- Les gains en termes de couverture ne compensent pas les pertes en termes de fréquence.

TABLE 6

DETERMINATION DU DOSAGE : CONCENTRATION - DIVERSIFICATION

AVANTAGES DE LA DIVERSIFICATION

- Profiter des effets synergiques.
- Bénéficier des complémentarités entre media.
- Obtenir une couverture nette plus élevée.
- Diminuer la variance du nombre d'expositions.
- Améliorer la répartition géographique.
- Minimiser la confrontation directe avec la concurrence.

AVANTAGES DE LA CONCENTRATION

- Dominer le media le plus adapté à la cible.
- Personnaliser la campagne et le produit.
- Choisir le support le plus adapté au produit.
- Bénéficier d'économies d'échelle, de ristournes.
- Minimiser le gaspillage.
- Attaquer frontalement la concurrence.

INCONVENIENTS DE LA DIVERSIFICATION

- Risque de dilution des efforts.
- Budget important.
- Moindre sélectivité.

INCONVENIENTS DE LA CONCENTRATION

- Moindre flexibilité en cas d'erreur de tir ou de riposte massive de la concurrence.
- Non couverture de certains segments du marché et de certaines régions géographiques.
- Couverture nette plus faible.
- Variance élevée du nombre d'expositions.

dans sa description de la réalité, ne peut que désigner le meilleur parmi l'ensemble des plans qui lui sont soumis. Si bien que l'on n'a guère le choix qu'entre la solution optimale d'une représentation appauvrie du problème et une solution non optimale d'une représentation plus fidèle.

Selon leur philosophie, les modèles de planification publicitaire utilisent l'une des quatre techniques suivantes (2) : la tabulation, la programmation mathématique sous ses diverses formes, la programmation heuristique et la simulation.

La technique de tabulation consiste en la construction de tableaux croisés, classifiant les individus-cibles selon leur degré d'exposition aux divers supports sélectionnés d'une part, et selon leurs caractéristiques socio-démographiques et psychographiques ou leur comportement d'achat d'autre part. La sélection des supports à tabuler, tout comme l'évaluation et la comparaison des performances de chaque combinaison sélectionnée, demeure entièrement soumise à la synthèse personnelle du spécialiste media. L'introduction explicite de coefficients de pondération assignés à chacun des supports et à chacun des segments de la cible ne modifie pas fondamentalement le caractère subjectif de cette approche (3) (4). Toutefois, la méthode possède une grande souplesse d'application et permet de tirer le meilleur parti possible de l'expérience accumulée au cours de campagnes similaires.

Exclusivement utilisée avant l'introduction de l'ordinateur, puis un moment éclipsée par l'engouement pour les méthodes optimisantes, elle connaît actuellement un regain de faveur causé par la mise au point de programmes de tabulation conversationnels tels que Crosstab (5) et Fastab (6) aux USA et Criven (7) en Belgique. Ces programmes, extrêmement versatiles grâce aux mémoires magnétiques à accès rapide, permettent de croiser ensemble un grand nombre de critères et de parvenir ainsi à une définition extrêmement précise de la cible. Certains d'entre eux accomodent l'introduction de pondération et permettent de classer les supports selon des échelles de puissance, d'économie et de sélectivité.

Les principaux avantages de la technique de tabulation sont sa facilité d'utilisation et son faible coût. Elle ne constitue, toutefois, guère plus qu'une méthode d'organisation des données disponibles.

Les techniques de programmation mathématique sont plus ambitieuses puisqu'elles se proposent de déterminer la combinaison de supports qui maximise

TABLE 7

LES MODELES DE PLANIFICATION PUBLICITAIRE : UNE VUE D'ENSEMBLE

3 FAMILLES	Modèlesde sélection des supports. Modèlesde chronologie des insertions. Modèlesde planification publicitaire.
2 PHILOSOPHIE	Optimisation. Evaluation.
4 TECHNIQUES	Tabulation. Programmation mathématique. Programmation heuristique. Simulation.
7 DOMAINES	Les facteurs d'audience. Les facteurs d'exposition. Les facteurs de média. Les facteurs de message. Les facteurs de fréquence. Les facteurs géographiques. Les facteurs concurrentiels.
7 CRITERES	L'exposition aux supports. L'exposition pondérée aux supports. L'exposition aux messages. Le souvenir, la réminiscence spontanée. Les modifications d'attitude. Les intentions d'achat. Le comportement d'achat.

un critère d'efficacité donné, en général unique. On en distingue quatre : la programmation linéaire, la programmation non linéaire, la programmation dynamique et la méthode itérative. Les principales faiblesses des méthodes de programmation mathématique sont présentées dans la table 8. Les tables suivantes présentent la programmation linéaire (table 9), la méthode itérative High Assay (table 10) et la méthode heuristique de Little et Lodish (table 11).

Outre leur famille, leur philosophie et leur technique, les modèles de planification publicitaire varient en matière de domaine d'application. Le domaine d'application d'un modèle s'apprécie à la place faite aux facteurs d'audience, d'exposition, de media, de message, de fréquence et de concurrence dans la détermination ou l'évaluation du plan publicitaire.

Les facteurs d'audience peuvent être classés en cinq catégories :

a) Les décompositions socio-démographiques et psychographiques qui répondent à la question : qui fait partie de l'audience du support ?

b) Les décompositions géographiques qui nous apprennent où se situe l'audience d'un support ;

c) Les décompositions temporelles selon l'heure du jour ou l'époque de l'année, qui précisent les fluctuations de l'audience à travers le temps ;

d) Les décompositions liées au produit, qui divisent l'audience en grands, moyens et petits consommateurs d'un produit particulier ou d'une marque particulière ;

e) Les décompositions liées aux supports qui décrivent la distribution de l'audience par quintiles, allant des téléspectateurs les plus assidus aux téléspectateurs les plus occasionnels, ou des lecteurs primaires par opposition aux lecteurs occasionnels, ou des abonnés par opposition aux acheteurs au numéro ... et ainsi de suite.

Les facteurs d'audience sont particulièrement utiles pour décrire et identifier la population-cible visée par la campagne publicitaire (8). Un modèle publicitaire satisfaisant doit pouvoir tenir compte de pondérations de la population-cible, qui indiquent l'importance relative pour l'annonceur des différents segments de la population totale.

TABLE 8

PRINCIPALES FAIBLESSES DES METHODES DE PROGRAMMATION MATHEMATIQUE

A - PROGRAMMATION LINEAIRE

1 - L'hypothèse de la linéarité de la fonction objectif et des contraintes ne correspond pas à la réalité. Elle implique que 2 individus exposés 1 fois soient considérés équivalents à 1 individu exposé 2 fois !

2 - L'accumulation et la duplication des audiences ne sont pas considérées.

3 - Les coûts sont supposés proportionnels au nombre d'insertions. Les ristournes ne sont pas considérées.

4 - La chronologie des insertions n'est pas considérée.

5 - Les facteurs liés au message ne sont pas considérés.

6 - Les pondérations d'audience et de support sont laissées à l'arbitraire de l'utilisateur de même que la fixation du nombre maximum d'insertions par support.

7 - La programmation en nombre entier est coûteuse.

B - PROGRAMMATION NON LINEAIRE

1 - L'hypothèse de la convexité de l'espace des solutions n'est pas vérifiée à cause de la discontinuité des coûts d'achat d'espace.

2 - La chronologie des insertions n'est pas considérée.

3 - L'optimisation est très coûteuse.

C - PROGRAMMATION DYNAMIQUE

1 - Ne peut résoudre que des problèmes de petite taille (moins de 15 supports).

2 - Coût prohibitif.

D - METHODE ITERATIVE

1 - Ne garantit pas une solution optimale.

2 - Ne considère pas la chronologie des insertions (sauf High Assay).

TABLE 9

LES METHODES D'OPTIMISATION

A - LA PROGRAMMATION LINEAIRE

FONCTION OBJECTIF : MAXIMISER LE NOMBRE D'EXPOSITIONS PONDEREES

Maximise : $E = e_1 X_1 + e_2 X_2 + \ldots\ldots + e_n X_n$

CONTRAINTES :

1. De budget $c_1 X_1 + c_2 X_2 + \ldots\ldots\ldots\ldots + c_n X_n \leqslant B$

2. De répartition du budget par média

$$c_1 X_1 + c_2 X_2 \leqslant B_1$$

$$c_3 X_3 + c_n X_n \leqslant B_2$$

3. Maximum d'insertions

$$X_1 \leqslant M_1$$

$$X_2 \leqslant M_2$$

$$\vdots \quad \vdots$$

$$X_n \leqslant M_n$$

4. Minimum d'insertions

$$X_1 \geqslant m_1$$

$$X_2 \geqslant m_2$$

$$\vdots \quad \vdots$$

$$X_n \geqslant m_n$$

SYMBOLES :

E = nombre d'expositions pondérées (dupliquées ou brutes)
e_i = efficacité d'une insertion dans le support i
X_i = nombre d'insertions dans le support i
c_i = coût d'une insertion dans le support i
B = budget disponible
B_1 = part du budget devant être alloué au média 1 (supports 1 et 2)
B_2 = part du budget devant être alloué au média 2 (supports 2 à n)
M_i = maximum d'insertions permises dans le support i
m_i = minimum d'insertions dans le support i

TABLE 10

LES METHODES D'OPTIMISATION

B - LA METHODE INTERATIVE HIGH ASSAY

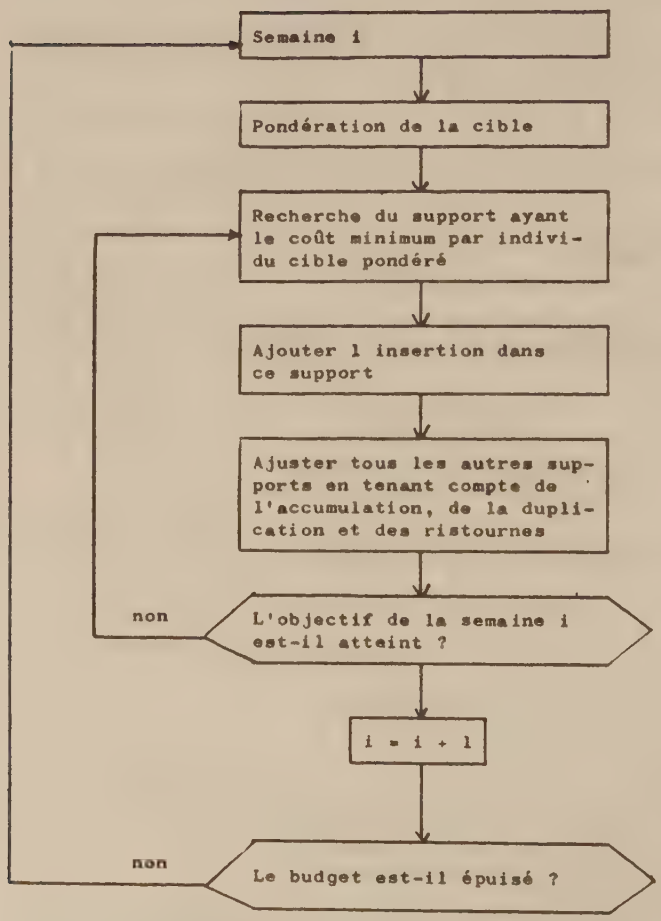
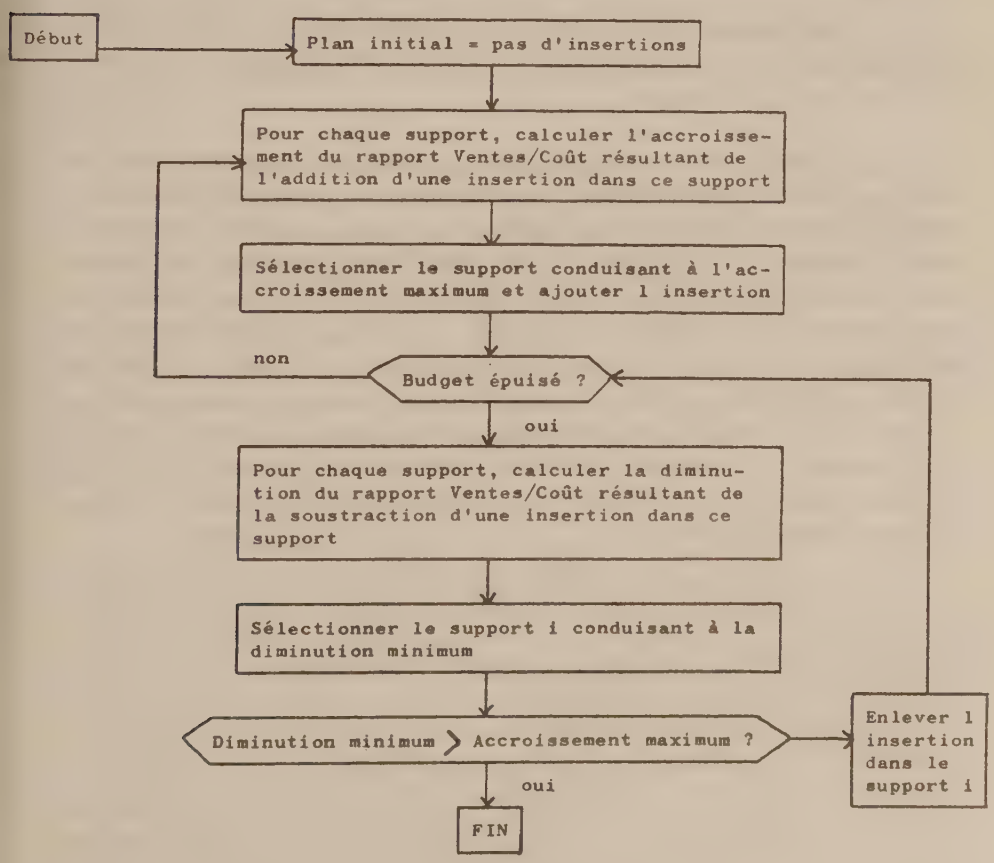


TABLE 11

LES METHODES D'OPTIMISATION

C - LA METHODE HEURISTIQUE DE LITTLE ET LODISH



Les facteurs d'exposition peuvent être divisés en deux catégories : exposition au support publicitaire et exposition au message publicitaire, une fois que l'exposition au support a effectivement eu lieu.

Le premier problème traite de la prédiction du nombre d'individus-cibles qui auront l'occasion de voir ou d'entendre un support publicitaire. La connaissance du nombre total d'individus exposés présente un intérêt évident, mais plus importante encore est la prédiction de la fréquence d'exposition parmi les individus-cibles, car certains d'entre eux seront exposés de nombreuses fois, alors que d'autres ne le seront que quelques fois.

C'est ce problème qui est le sujet de notre thèse et les quatre paragraphes suivants de ce chapitre décrivent les différentes approches qui en ont été faites par la littérature.

Un modèle satisfaisant d'exposition aux supports doit fournir des procédés par lesquels la distribution des occasions de voir peut être évaluée pour n'importe quel plan d'insertions publicitaires. De plus, il devrait aussi permettre l'estimation de la distribution des expositions aux messages publicitaires, étant donné la distribution des expositions aux supports. En ce qui concerne les media imprimés, ceci peut être réalisé grâce au concept de l'exposition à la page publicitaire dont Politz (9) fut le pionnier, complété par le nombre de jours de lecture, la durée des lectures et la proportion de numéros lus, fournis par Simmons (10). Pour ce qui concerne les media radiodiffusés, les échelles d'attention rapportées par Simmons et les indices d'écoute (T.V.Q.) produit par la "Home Testing Institute" peuvent être utilisés pour réduire proportionnellement les audiences brutes de programmes.

Les facteurs tenant aux media évaluent l'aptitude d'un support donné à transmettre un message particulier. A titre de comparaison, on peut dire que, si les facteurs d'audience permettent d'harmoniser les supports avec le marché, les facteurs de media permettent d'harmoniser les supports avec le message. L'action du support ne se limite pas à convoyer le message publicitaire. Par le jeu des inévitables connotations liées à la personnalité même du support, celui-ci contribue, positivement ou négativement, à l'efficacité finale du message, en créant une disposition, un état d'esprit, en faisant naître la confiance, en suggérant la qualité ou le prestige, ou bien, tout simplement, parce que ses lecteurs s'attendent à ce qu'il transmette des messages publicitaires sur certaines catégories de produits plutôt que sur d'autres. Un grand nombre d'études sur l'image des media se sont développées pendant les années cinquante et le début des années

soixante, et un bon compte rendu en est fournj dans le chapitre XV du livre de Lucas et Britt (11). Plus récemment, les études de Banks (12), Swanson (13), Frank, Beckwell et Clokey (14), Tigert (15), Wells, Banks et Tigert (16) et Wells (17), ont tenté de démontrer que les programmes télévisés forment des groupes distincts, auxquels il est possible d'attribuer diverses pondérations en fonction des objectifs de l'annonceur. Des études similaires, effectuées par Wilson (18), Bass, Pessemier et Tigert (19), Bernay (20), King et Summers (21), Tigert (22), font apparaître des conclusions identiques pour les media imprimés.

Un modèle satisfaisant de planification publicitaire se doit donc d'inclure des facteurs de media afin de résumer les effets du climat éditorial, de l'affinité du support avec la nature du produit annoncé, du volume de publicité placé par les concurrents, du format du support et de la qualité technique de ses reproductions.

Les facteurs liés au message comprennent toutes les caractéristiques, tenant tant à la forme qu'au contenu du message publicitaire, qui influencent la perception, la mémorisation, la compréhension et l'évaluation du message par l'audience-cible. Un modèle satisfaisant de planification publicitaire se doit donc de comporter des sous-programmes spécialisés, chargés de prédire le degré de perception, de compréhension, de mémorisation et d'acceptation du message publicitaire. Par exemple, en ce qui concerne les media imprimés, la probabilité conditionnelle de perception, l'exposition ayant eu lieu, est fonction de la taille et des couleurs du message, de la position de la page à l'intérieur du numéro, et de l'emplacement du message dans la page ainsi que de l'épaisseur du numéro et de l'intérêt que présente pour le lecteur le produit annoncé, comme cela a été démontré dans différentes études consacrées à la perception des messages publicitaires (23), (24), (25). Les effets liés au contenu même du message sont beaucoup moins connus, quoique les études sémiologiques commencent à se développer (26). A titre d'exemple, les deux questions suivantes sont encore très controversées : faut-il user d'arguments de vente agressifs, ou bien convient-il d'utiliser des arguments plus feutrés ? L'argumentation doit-elle se contenter de présenter les avantages du produit, ou bien doit-elle, en outre, tenter de réfuter les objections probables des lecteurs ? On peut trouver un bon exposé des contributions apportées par la recherche psychologique à l'étude des stratégies de communication dans le livre de Léo Bogart (27). L'examen critique le plus récent traitant des réactions cognitives à la publicité peut être trouvé dans la recherche de Wright (28) qui conclut tristement que l'état présent des théories des effets de la publicité est équivoque et frustrant par son degré d'inorganisa-

TABLE 12

PONDERATION DES MESSAGES

FACTEURS DE FORMAT

PONDERATION DE LA TAILLE - PUBLICITE NOIR ET BLANC

TAILLE (en pages)	.25	.33	.50	.66	1	2	3	4
PONDERATION	41	47	58	69	91	156	221	285

PUBLICITE EN COULEUR

TAILLE (en pages)	.33	.50	.66	1	2
PONDERATION 2 couleurs	35	52	67	97	175
PONDERATION 4 couleurs		103		134.	186

NOMBRE DE MOTS DANS LE TITRE

NOMBRE DE MOTS	0	1 à 3	4 à 6	7 à 9	10 ou plus
PONDERATION	79	101	99	96	94

TYPE D'ILLUSTRATION

		Nombre d'éléments	Pondération
PHOTOGRAPHIE D'UN INDIVIDU	120		
PHOTOGRAPHIE D'UN OBJET	100	4 au moins	100
DESSIN (Half-tone)	80	4 à 8	75
TABLEAU OU GRAPHIQUE	90	8 ou plus	48

STYLE PUBLICITAIRE

	Attention	Intérêt	Rétention
STYLE EDITORIAL	82	78	156
DOMINANCE ILLUSTRATION	112	112	116
STYLE BANDE DESSINEE	95	92	290

Source : E.P. Schonfeld, "Prediction of Readership and Enquiry Response to Print Advertising. Ph. D. Thesis, Northwestern University, August 1975.

TABLE 13

PONDERATION DES MESSAGES

B - FACTEURS DE POSITION

POTISITION	PONDERATION
COUVERTURE	174
VERSO COUVERTURE	131
DOS	160
VERSO DOS	124
PAGES INTERIEURES	100

EPAISSEUR DU MAGASINE

NOMBRE DE PAGES	PONDERATION
MOINS DE 100	100
100 - 200	88
120 - 140	82
140 - 160	77
160 - 180	75
180 - ET PLUS	73

Source : E.P. Schonfeld, "Prediction of Redearship and Enquiry Response to Print Advertising", Ph. D. Thesis, Northwestern University, August 1975.

tion. Sur le plan pratique, des conseils avisés sur la manière d'introduire les facteurs tenant au message dans un modèle de planification publicitaire sont fournis par Maloney (29), (30) et Axelrod (31), alors que Gross (32) présente un modèle de production et de filtrage des thèmes des messages.

Les facteurs liés à la fréquence d'exposition ont trait à la question suivante : combien de répétitions sont nécessaires pour atteindre un objectif de communication particulier ? Une grande quantité de recherches a été publiée dans le domaine depuis les premières études d'Ebbinghauss (33) et de Strong (34), si bien qu'il est difficile de résumer leurs conclusions en quelques phrases. Néanmoins, si l'on prend le souvenir comme critère d'efficacité de la communication, on peut affirmer qu'aucune impression durable n'est acquise à la suite d'une seule exposition, que les expositions successives accroissent le degré de rétention mais à un taux décroissant, que le taux d'oubli est très rapide après l'exposition et qu'il se nivelle ensuite graduellement, et que plus le message est chargé de signification pour la personne exposée, moins elle l'oubliera. Le taux d'oubli dépend des activités qui prennent place entre les expositions, spécialement l'exposition à des messages concurrentiels, et des activités qui adviennent pendant l'exposition, spécialement les facteurs de distraction. L'importance des interférences proactives et des interférences rétroactives est fonction du degré de similarité existant entre le message et les messages interférents. L'application de ces résultats à la détermination de la chronologie d'une campagne publicitaire est habilement résumée par Durand (35). Deux modèles de fixation de la chronologie des insertions, basés sur les résultats de l'étude de Zielske (36), ont été présentés par Huges (37) et par Strong (38). Un modèle similaire est actuellement employé par l'agence Léo Burnett (39).

Quand on s'intéresse à d'autres critères d'efficacité de la communication tels que les croyances, les comportements, les intentions d'achat, le nombre de coupons retournés, ou même le nombre d'achats effectués, les effets de la répétition sont moins clairs. Un certain nombre d'études de laboratoire sur les effets de la répétition par Ray et Sawyer (40), (41), Ray, Sawyer et Strong (42) ont fait apparaître des différences d'effet dramatiques selon la nature de la situation publicitaire.

Le niveau, la forme et la pente des fonctions de réponse observées semblent dépendre de la nature des réactions mesurées, de la nature du produit, du contenu du message publicitaire, de la familiarité du répondant avec le produit, du thème utilisé, du volume de messages concurrentiels et des media utilisés,

accumulant la distribution des occasions de voir en partant du nombre d'individus exposés exactement une fois. La réponse à la deuxième question s'obtient en décumulant la distribution en partant de la couverture nette, ainsi que l'illustre le tableau suivant .

Nombre d'Expositions	Distribution Cumulée	Distribution Décumulée
1	11,7 %	28,6 %
2	23,7 %	16, 9 %
3	26,8 %	4,9 %
4	28,6 %	1,8 %

Lorsque l'on désire s'en tenir à une comparaison rapide de deux distributions, on calcule la fréquence moyenne d'exposition en divisant l'audience brute par la couverture nette. L'audience brute d'un plan publicitaire est égale à la somme des audiences brutes de chacun des supports.Pour notre exemple nous obtenons :

Audience brute = 20 832 * 2 + 13 984 * 2 = 69 632

Couverture nette = 29 216

Fréquence moyenne = 2,38

On peut noter qu'il est possible de calculer l'audience brute à partir de la distribution des occasions de voir en multipliant l'effectif de chaque classe par le nombre d'expositions correspondant et en faisant la somme pour l'ensemble des classes d'exposition. Ceci montre que l'audience brute n'est autre que le nombre total d'expositions générées par le plan publicitaire.

Audience brute = 15 569 * 1 + 16 062 * 2 + 4 193 * 3 + 2 340 * 4 = 69 632

Finalement, un dernier aspect du caractère dynamique des audiences est décrit par les concepts d'audiences fidèles et d'audiences infidèles.

10. Audiences fidèles et Audiences Infidèles

A mesure que le temps s'écoule, le nombre d'individus exposés à toutes les parutions successives d'un support diminue. Plus cette décroissance est lente, plus l'audience est loyale. En conséquence, l'audience loyale d'un support peut être définie par le nombre d'individus exposés à N parutions parmi N.

Ainsi, quel que soit le nombre de parutions considérées, la distribution des audiences entre les diverses classes de fidélité résume parfaitement la capacité du support à maintenir une audience captive.

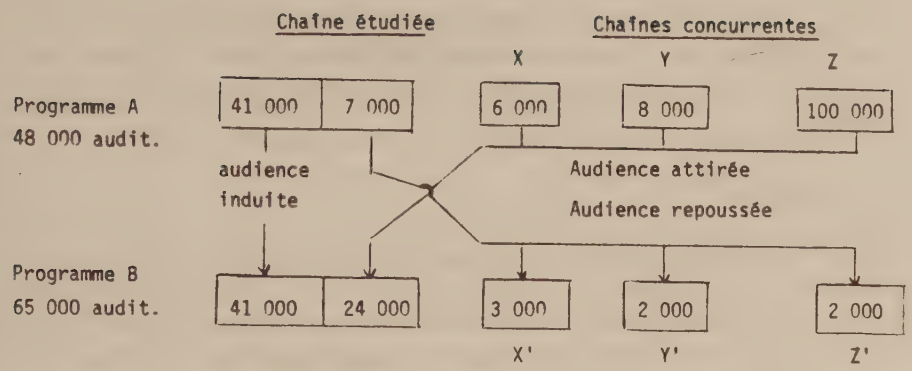
A l'opposé, l'infidélité d'une audience peut être mesurée par le nombre d'individus qui quittent ou entrent dans l'audience d'un support pendant une période donnée. Si l'on considère, par exemple, une chaîne de télévision, trois questions se posent quant à la fidélité des audiences d'un programme particulier.

a- Combien d'individus faisant partie de l'audience du programme A feront encore partie de l'audience du programme suivant B, présenté sur la même chaîne ? Cette audience fidèle à la chaine s'appelle l'audience induite (50). L'audience induite d'un programme quelconque est toujours inférieure à l'audience globale de ce programme car il y a toujours un certain pourcentage d'auditeurs infidèles qui se dirigeront vers une autre chaîne à la fin du programme A.

b- Combien d'individus ne faisant pas partie de l'audience du programme A feront partie de l'audience du programme B, et d'où viennent ces nouveaux auditeurs ? Cette audience nouvellement acquise à la chaîne est l'audience attirée (50) par le programme B. Elle mesure la capacité de séduction du programme B auprès des auditeurs des chaînes concurrentes et auprès des non-auditeurs.

c- Combien d'individus faisant partie de l'audience du programme A ne feront plus partie de l'audience du programme B et où sont-ils allés ? Cette audience perdue par la chaîne est l'audience repoussée (50) par le programme B. Elle mesure à la fois l'incapacité du programme B à conserver certains auditeurs du programme précédent et l'attirance combinée exercée par les programmes concurrents.

Un bon programme se doit donc de conserver les auditeurs du programme précédent, de séduire les auditeurs des programmes concurrents et de léguer un maximum d'auditeurs fidèles au programme suivant. Le diagramme ci-après illustre les trois concepts d'audience induite, attirée et repoussée.



Une description complète de la concurrence entre les chaînes est donnée par la matrice des changements de chaîne qui décrit les mouvements d'audience d'un programme à un autre.

		Des Programmes				
		A	X	Y	Z	Audience des programmes d'arrivée
Vers les Programmes	B	41 000	6 000	8 000	10 000	65 000
	X'	3 000	4 000	1 000	3 000	11 000
	Y'	2 000	11 000	6 000	2 000	21 000
	Z'	2 000	9 000	1 000	3 000	15 000
Audience des Programmes de départ		48 000	30 000	16 000	18 000	

Cette matrice permet de calculer, pour chacun des programmes en cause, les audiences induites, attirées et repoussées

Programmes	Audiences induites	Audiences attirées	Audiences repoussées
A	41 000	24 000	7 000
X	4 000	7 000	26 000
Y	6 000	15 000	10 000
Z	3 000	12 000	15 000

Nous avons maintenant terminé notre effort de définition des principaux concepts et notre analyse des divers problèmes de mesure y afférant. Les principales conclusions qui se dégagent de cette étude préliminaire sont les suivantes .

11. Conclusions

En ce qui concerne la mesure des audiences des media imprimés nous avons mis à jour les divers biais qui accompagnent le choix d'une technique de collecte des données, d'un critère de lecture et d'une formulation des questions. Ces divers biais ont été illustrés par des données empiriques, tant françaises qu'américaines, attestant de l'ampleur des désaccords entre les diverses méthodologies.

Une étude similaire a été conduite pour les media radiodiffusés. Les inconvénients et les limitations des diverses procédures utilisées par Nielsen, ARB et Simmons ont été détaillés et illustrés par la comparaison des audiences estimées pour divers programmes télévisés. De même, les méthodologies utilisées dans le domaine des audiences radio par Arbitron, Hooper, Market Trends, Pulse, The Source et Trendex ont été présentées et comparées.

Une énumération extensive des différentes classifications d'audience a été compilée, accompagnée d'une description des techniques informatiques de tabulation multicritère. Les divers concepts qui permettent de définir et de mesurer le caractère dynamique des audiences ont été analysés. Les concepts nouveaux de matrice d'exposition et de distribution des audiences fidèles ont été proposés afin de systématiser et de compléter la compréhension des phénomènes d'exposition aux supports publicitaires.

Malgré la controverse actuelle concernant la qualité des données d'audience (51), (52), controverse mise à feu par les divergences croissantes séparant les estimations de Simmons et celles de TGI (53), (54),il faut bien reconnaître que l'on dispose aujourd'hui de données d'audience incomparablement plus complètes et plus détaillées que jamais dans le passé (55). Les insatisfactions récentes peuvent être attribuées davantage à une réalisation soudaine des difficultés de la mesure des audiences qu'à une détérioration de la qualité des recherches.

NOTES

(1) - GRAHAM I., "Encyclopedia of Advertising", New York, 2d Edition, 1969, Fairchild.

(2) - Standard Rate & Data Service, "Dictionary of Termes Useful to Buyers of Advertising", 1959, Skokie Ill.

(3) - Advertising Research Foundation, Audience Concept Committee, "Towards Better Media Comparisons", New York, 1961

(4) - BROWN I.O, LESSLER R.S, WEILBACHER, W.M., "Advertising Media : Creative Planning in Media Selection", New York, 1967, Ronald Press, Ch. 4.

(5) - SANDAGE C.H., FRYBURGER V. , "Advertising Theory and Practice", Ch. 19, 8 th Edition, 1971, Irwin.

(6) - LUCAS D.B., BRITT, S.H., "Measuring Advertising Effectiveness", ch. 9, Mc Graw Hill, 1963.

(7) - Cette définition suit de très près celle proposée par l'A.R.F. Audience Concept Committee, et dont l'auteur est M. Paul E.J. Gerhold.

(8) - Audit Bureau of Circulation, 123 North Wacker Drive Chicago, Illinois.

(9) - Standard Rate & Data Service, Inc. 5201 Ol d Orchard Road, Skokie, Illinois

(10) - A.C. NIELSEN Co., Northbrook, Illinois

(11) - Le compteur appelé "audimeter", enregistre sur un film le canal sur lequel le récepteur est branché.

(12) - American Research Bureau Inc., 1350 Avenue of America , New York 10019

(13) - Television Information Office, How Good are Television Ratings ? T.I.O., 66 Fifth Avenue, New York

(14) - La différence entre la pénétration et la circulation d'un support imprimé est probablement très faible puisque très peu d'individus omettent de lire une publication qu'ils viennent d'acheter.

(15) - Cette affirmation doit être tempérée pour tenir cas d'une possible destruction de l'affiche par lacération ou intempéries.

(16) - Centre d'Etudes des Supports de Publicité, 8 Rue Daniel Casanova Paris 2ème.

(17) - W.R. SIMMONS & Assocociates Research, Inc. 235 East 42 nd Street, New York, N.Y. 10017

(18) - SECODIP, Rue Francis Pedron, 78241 Chambourcy

(19) - Target Group Index, American Market Research Bureau, New York 420 Lexington Avenue N.Y. 100017

(20) - Starch D. & Staff Inc, Starch Consumer Magazine Reports

(21) - Alfred POLITZ Research, Inc, New York. La méthode a été inventée par Louise Mc Carty Dubois travaillant pour le magazine Life.

(22) - T. JOYCE, "A Comparison of methods", Axiom Market Research Bureau Nov. 1973

(23) - Daniel MUGNIER, "Recherche d'une formulation optimale des questions d'habitudes de lectures et de lecture dernière période dans les mesures d'audience des magazines", 12° journées d'études de l'IREP, octobre 1972.

(24) - CORLETT T. & OSBORNE D.W. "The development of reading frequency scales" IPA, London 1966

(25) - DURAND J. "Recherche des données nécessaires au fonctionnement des modèles", la Recherche opérationnelle appliquée à l'élaboration des plans media, IREP, 1966

(26) - GREEN J.D. "Personal Media Probabilities", Journal of Advertising Research, October 1970, pp. 12-18

(27) - SCHREIBER R.J. "Probability Assignments for the simulation of Media Reach and Frequency", Journal of Advertising Research, June 1968

(28) - Alfred Politz Media Studies, "An Experimental study comparing magazines audiences as determined by two questioning procedures". ARF Conference 1967

(29) - SIMMONS W.R. "A new look at Reach and Frequency", ARF Conference 1969

(30) - PARFITT J.H. "The Use of Panels for the collection of Readership and others Media Data" The Thomson Medal, 1966 pp. 73-109

(31) - LANCESTRE R., PINET B., "Contribution of consumer Panels to the Study and Use of Periodicals by Media-Planners", Marcel Dassault Award, 1970, pp. 41-71

(32) - JOYCE T., BIRD M., "The use of panels in readership surveys", The Thomson Awards for Media research, 1966, pp. 49-70

(33) - PINET Bernard, DERIBERE-DESGARDES Pierre, "La Contribution des Panels de Consommateurs et des Panels Spécifiques à la mesure des Audiences des Supports de Publicité" Etude N°27, IREP, Juillet 1973

(34) - Newspaper Advertising Bureau Inc. 485 Lexington Avenue, New York

(35) - L'Aurore, La Croix, L'Equipe, Le Figaro, France-Soir, L'Humanité, Le Monde, Le Parisien Libéré

(36) - ALLEN C.L. "Photographing the TV Audience" Journal of Advertising Research, March 1965, pp. 2-8

(37) - L'appellation originale des 5 index est la suivante
1 = Coverage, 2 = Honseholds Using Television (HUT),
3 = Total Audience, 4 = Average Audience
5 = Share of Audience

(38) - Traduction du mot "rating"

(39) - American Research Bureau Inc.

(40) - Traduction des termes "coincidental interview" et "retrieval interview"

(41) - Pulse, Inc. 730 Fifth Avenue, New York

(42) - Hooper C.E., Mamaroneck, New York

(43) - Market Trends, 100 Wilmot Road, Deerfield, Illinois

(44) - Trendex, 509 Madison Avenue, New York

(45) - Dimension Unlimited, 2816 Rowena, Los Angeles

(46) - The Value in Arbitron Radio, 1974, American Research Bureau, Inc.

(47) - Cette catégorie inclue les publicités imprimées au dos d'enveloppes de lettres et les tracts publicitaires

(48) - Traduction du terme "Gross Audience"

(49) - J.W. Thompson Agency, The concept of effective reach, internal memo, Novembre 1973, New York

(50) - Traduction des termes "Lead in" , "switch in" , switch out"

(51) - "Roundtable 7 : The Credibility Gap in Media Research", Media Decisions, April 1974

(52) - "Magazine Research : Who is suing Who ?" Media Decisions, February 1975

(53) - Papazian, E., "Media Research Jumps from the Frying Pan into the Fire", Media Decisions, December 1974, pp. 12-14

(54) - "Agency Researchers Stumped by Big Swing in New Simmons Numbers", Media Decisions, December 1974, pp. 28-29

(55) - A.R.F.'s Total Audit of Simmons, A.R.F., Vol. I, II, III, IV, New York, August 1975

CHAPITRE 5

LES MODELES D'ACCUMULATION D'AUDIENCE

Tout au long de ce chapître, notre analyse sera centrée sur l'étude des plans d'insertions qui font appel à un support publicitaire unique. Ce type de plan publicitaire n'est certes pas fréquent dans la pratique. Néanmoins, la compréhension des modèles d'accumulation d'audience constitue la base nécessaire à la construction des modèles d'exposition plus complexes qui font intervenir plusieurs insertions dans plusieurs supports.

L'objet de tout modèle d'accumulation d'audience est d'estimer la couverture nette et la distribution des occasions de voir engendrée par N insertions publicitaires placées dans N parutions différentes, consécutives ou non, d'un même support publicitaire. En conséquence, il est naturel de commencer l'analyse par l'étude de la relation qui lie l'accumulation de l'audience à la distribution des occasions de voir.

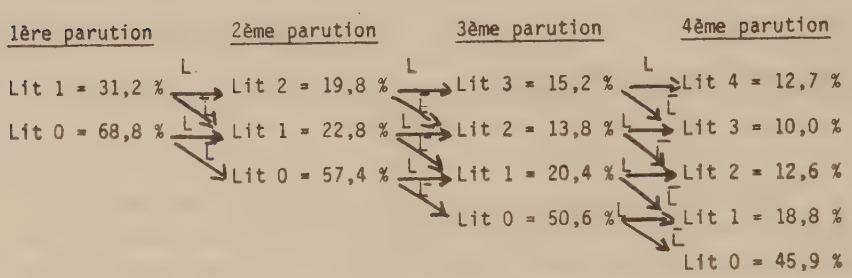
1. - La relation fondamentale liant l'accumulation de l'audience à la distribution des occasions de voir.

Le mécanisme qui lie la couverture nette à la fréquence d'exposition lorsque les parutions successives d'un support publicitaire sont prises en considération s'expose le mieux en commençant par un exemple : l'étude de l'audience du magazine "Life", effectuée par Politz (1) en 1958 nous apprend que 31,2 % des foyers américains étaient couverts, en moyenne, par un numéro, 42,6 % étaient couverts par deux numéros, 49,4 % par trois numéros et 54,1 % par quatre numéros. Cette étude nous apprend aussi qu'à la suite de quatre parutions, 45,9 % des foyers n'avaient été exposés à aucun des quatre numéros, 18,8 % avaient été exposés une fois seulement, 12,6 % deux fois, 10,0 % trois fois et 12,7 % avaient

été exposés aux quatre numéros. La table suivante décrit la relation qui lie ces deux ensembles de nombres.

Accumulation de l'audience du magazine "Life" dans les foyers américains en 1958

(L signifie Lecture ; $\bar{L}$ signifie non Lecture)

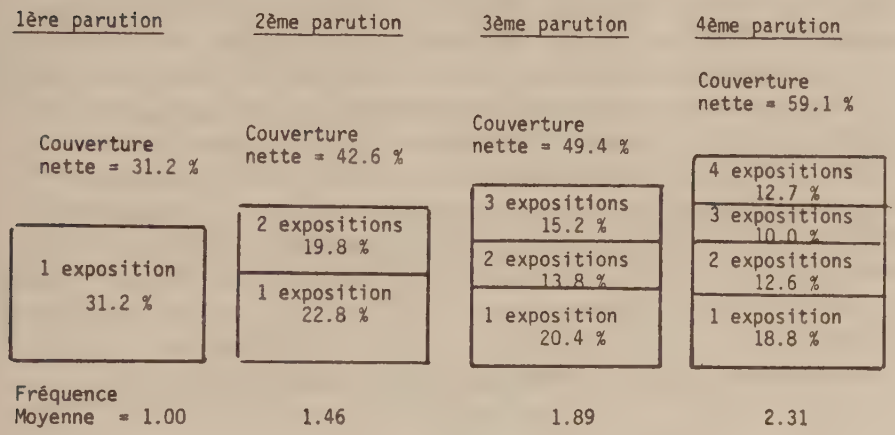


Couverture nette cumulée	= 31,2 %	42,6 %	49,4 %	59,1 %
Fréquence moyenne	= 1,00	1.46	1.89	2.31

A la lecture de cette table, la relation fonctionnelle qui lie la couverture nette cumulée à la distribution des occasions de voir devient claire. A la première parution, la couverture nette est égale au pourcentage d'individus exposés une fois. A la seconde parution, la couverture nette cumulée est égale au pourcentage d'individus exposés à chacune des deux parutions plus le pourcentage d'individus exposés à la première parution mais pas à la seconde, et plus le pourcentage d'individus exposés à la seconde parution mais pas à la première, (et ainsi de suite pour les parutions suivantes). Ainsi, le taux d'accumulation de l'audience d'une parution à une autre modèle la distribution des occasions de voir qui, elle-même, reflète toutes, les accroissements successifs de la couverture nette.

Ce phénomène peut être rendu encore plus explicite à l'aide du graphique suivant.

Accumulation de l'audience du magazine "Life" dans les foyers américains en 1958.



Nous allons démontrer que, sous l'hypothèse de la constance de l'audience moyenne d'un numéro, la distribution des occasions de voir peut être déduite mathématiquement de la connaissance de la couverture nette cumulée et réciproquement. Ce fait avait déjà été relevé par J.D. Greene et J.S. Stock en 1967 (2), dans leur brochure "Advertising Reach and Frequency in Magazines", mais sans que soit fournie la démonstration ni la discussion de l'hypothèse sous-jacente (*)

12 - De la distribution des occasions de voir à la couverture cumulée.

Supposons que la distribution des occasions de voir générée par cinq parutions d'un support donné soit la suivante :

Fréquence d'exposition	0	1	2	3	4	5	Total
Nombre d'individus	50	20	10	5	5	10	100

(*) "La conversion de l'audience cumulée à N parutions à la distribution des occasions de voir est, en principe, une identité algébrique". Ouvrage cité.

Supposons, de plus, que l'exposition aux cinq parutions soit équiprobable. Equiprobable signifie que, quelle que soit la probabilité d'exposition d'un individu donné, cette probabilité ne dépend ni de la parution considérée ni de l'exposition aux parutions précédentes. Parmi les individus qui n'ont été exposés à aucun des cinq numéros, la probabilité d'exposition à n'importe lequel de ces cinq numéros est, bien entendu, nulle. Cependant, il convient de remarquer qu'elle ne serait pas nulle si nous considérions un sixième numéro. Parmi les individus qui ont été exposés à l'un quelconque des cinq numéros, quelle est la probabilité que cela fût au premier numéro ? L'exposition à un numéro sur cinq peut se produire de cinq manières différentes($\binom{5}{1} = 5$), chacune d'entre elles étantéquiprobable, la probabilité en question est de 1/5. Parmi les individus qui ont été exposés à deux numéros quelconques, la probabilité que l'un de ces deux numéros fût le premier est de $\binom{1}{1} * \binom{4}{1} / \binom{5}{2} = 2/5$. Plus généralement, la probabilité qu'un individu exposé à R numéros quelconques parmi N, ait été exposé au premier numéro est donnée par la loi hypergéométrique

$$\binom{1}{1} * \binom{N-1}{R-1} / \binom{N}{R} \quad \text{avec} \quad R \leqslant N.$$

Si nous nous demandons maintenant quelle est la probabilité d'exposition à la première parution pour l'ensemble de la population, nous pouvons trouver la réponse en étudiant chaque classe de distribution séparément et en les ajoutant, ainsi qu'on peut le voir dans le tableau suivant.

Calcul de la couverture nette de la première parution à partir de la distribution des occasions de voir

Fréquence	Individus	Probabilité d'exposition à la première parution	Individus prévus
0 sur 5	50	0	0
1 sur 5	20	1/5	4
2 sur 5	10	2/5	4
3 sur 5	5	3/5	3
4 sur 5	5	4/5	4
5 sur 5	10	1	10
Total	100		25

Le calcul de la couverture nette cumulée des deux premières parutions peut être obtenu également en utilisant la loi hypergéométrique. Par exemple :

la probabilité que les individus qui ont été exposés à une parution sur cinq aient été exposés à la 1ère ou à la seconde parution indifféremment, est donnée par :

Probabilité d'exposition à la première parution :

$$\binom{1}{1} * \binom{4}{0} / \binom{5}{1} = 1/5$$

Probabilité d'exposition à la deuxième parution :

$$\binom{1}{1} * \binom{4}{0} / \binom{5}{1} = 1/5$$

$$= 2/5$$

Cependant, il est plus rapide de calculer la probabilité de non exposition aux trois dernières parutions et puis de soustraire le résultat de 1 :

$$1 - \binom{2}{0} * \binom{3}{1} / \binom{5}{1} = 2/5$$

Calcul de la couverture nette cumulée des 2 premières parutions à partir de la distribution des occasions de voir

Fréquence	Individus	Probabilités	Individus prévus
0 sur 5	50	0	0
1 sur 5	20	2/5	8
2 sur 5	10	7/10	7
3 sur 5	5	9/10	4.5
4 sur 5	5	1	5
5 sur 5	10	1	10
Total	100		34.5

Le calcul de la couverture nette cumulée des trois premières parutions est traité de la même façon, et ainsi de suite jusqu'à la couverture nette cumulée des cinq parutions. Nous pouvons écrire une formule générale pour la conversion de la distribution des occasions de voir en couverture nette cumulée, après avoir adopté les notations suivantes :

N = Nombre de parutions pour lesquelles la distribution des occasions de voir est disponible.

K = Nombre d'expositions.

t = Nombre de parutions pour lesquelles on désire calculer la couverture nette.

R_t = Couverture nette cumulée après t parutions du support.

E_K^N = Nombre d'individus exposés à K parutions sur N.

$$(1)\quad R_t = \sum_{K=1}^{N} E_K^N \left[1 - \frac{\binom{N-t}{K}}{\binom{N}{K}} \right] \quad \text{avec } 0 > t \geqslant N \quad \text{et } K = 0, 1, 2, \ldots N$$

Notons que par convention $\binom{N-t}{K} = 0$ quand $K > N - t$

comme exemple de la façon dont la formule fonctionne, nous pouvons recalculer la couverture nette cumulée des deux premières parutions comme suit :

$$E_0^5 = 50,\ E_1^5 = 20,\ E_2^5 = 10,\ E_3^5 = 5,\ E_4^5 = 5,\ E_5^5 = 10$$

$$R_2 = E_1^5 \left[1 - \frac{\binom{3}{1}}{\binom{5}{1}} \right] + E_2^5 \left[1 - \frac{\binom{3}{2}}{\binom{5}{2}} \right] + E_3^5 \left[1 - \frac{\binom{3}{3}}{\binom{5}{5}} \right] + E_4^5 + E_5^5$$

$$R_2 = 20 * 2/5 + 10 * 7/10 + 5 * 9/10 + 5 + 10 = 34.5$$

Nous avons démontré que dans l'hypothèse de parutions équiprobables, la distribution des occasions de voir peut être convertie en couverture nette cumulée. Maintenant, nous allons prendre en considération le problème inverse.

13 - Conversion de la couverture nette cumulée en distribution des occasions de voir.

Supposons que la couverture nette cumulée générée par un plan publicitaire de 5 parutions dans un support donné soit comme suit :

Parutions	1	2	3	4	5	Population
Couverture nette cumulée	250	345	410	460	500	totale : 1000

Supposons, en outre, que les cinq parutions soient équiprobables. D'abord, nous pouvons aisément calculer le nombre d'individus qui ne sont exposés à aucune des cinq parutions =

$$E_0^5 = \text{population} - R_5 = 500$$

Nous pouvons aussi, en utilisant l'équation (1), écrire les cinq inconnues restantes, de E_1^5 à E_5^5 comme une fonction des cinq audiences cumulées connues.

Nous obtenons :

$$R_5 = E_1^5 + E_2^5 + E_3^5 + E_4^5 + E_5^5 = 500$$

$$R_4 = \frac{4}{5} E_1^5 + E_2^5 + E_3^5 + E_4^5 + E_5^5 = 460$$

$$R_3 = \frac{3}{5} E_1^5 + \frac{9}{10} E_2^5 + E_3^5 + E_4^5 + E_5^5 = 410$$

$$R_2 = \frac{2}{5} E_1^5 + \frac{7}{10} E_2^5 + E_3^5 + E_4^5 + E_5^5 = 345$$

$$R_1 = \frac{1}{5} E_1^5 + \frac{2}{5} E_2^5 + \frac{3}{5} E_3^5 + \frac{4}{5} E_4^5 + E_5^5 = 250$$

Nous avons maintenant un système de cinq équations avec cinq inconnues qui peut être résolu par l'inversion de la matrice des coefficients.

En réécrivant le système sous une forme matricielle, nous obtenons :

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4/5 & 1 & 1 & 1 & 1 \\ 3/5 & 9/10 & 1 & 1 & 1 \\ 2/5 & 7/10 & 9/10 & 1 & 1 \\ 1/5 & 2/5 & 3/5 & 4/5 & 1 \end{bmatrix} * \begin{bmatrix} E_1^5 \\ E_2^5 \\ E_3^5 \\ E_4^5 \\ E_5^5 \end{bmatrix} = \begin{bmatrix} 500 \\ 460 \\ 410 \\ 345 \\ 250 \end{bmatrix}$$

Soit A la matrice des coefficients, E le vecteur des fréquences et R le vecteur de la couverture nette cumulée, nous obtenons :

$$\underset{(5,5)}{A} * \underset{(5,1)}{E} = \underset{(5,1)}{R} ,$$

La solution de cette équation est :

$$E = A^{-1} * R$$

Mais les éléments de la matrice A sont fournis par l'équation (1), il est donc possible d'obtenir une formule générale pour les éléments de l'inverse

de la matrice A.

$$A^{-1} = (-1)^{i+j} * \begin{bmatrix} \binom{N}{0} & & & \\ & \binom{N}{1} & & \\ & & \ddots & \\ & & & \binom{N}{N} \end{bmatrix} * \begin{bmatrix} \binom{0}{0} & & & & \\ \binom{1}{0} & \binom{1}{1} & & & \\ \binom{2}{0} & \binom{2}{1} & \binom{2}{2} & & \\ \vdots & & & \ddots & \\ \binom{N}{0} & \cdots & \cdots & \cdots & \binom{N}{N} \end{bmatrix}$$

On obtient les coefficients hypergéométriques inverses en multipliant la matrice des coefficients binomiaux (matrice de Pascal) par une matrice diagonale de coefficients binomiaux, affectés du signe voulu. En appliquant ce resultat au cas de cinq parutions, nous obtenons :

$$E = \begin{bmatrix} 5 & -5 & & & \\ -10 & 20 & -10 & & \\ 10 & -30 & 30 & -10 & \\ -5 & 20 & -30 & 20 & -5 \\ 1 & -5 & 10 & -10 & 5 \end{bmatrix} * \begin{bmatrix} 500 \\ 460 \\ 410 \\ 345 \\ 250 \end{bmatrix} = \begin{bmatrix} 200 \\ 100 \\ 50 \\ 50 \\ 100 \end{bmatrix}$$

Le même résultat peut être réécrit sous une forme plus condensée :

(2) $$E_K^N = (-1)^K \binom{N}{K} \sum_{t=0}^{K} (-1)^{t+1} \binom{K}{t} R_{N-t}$$ où $K = 1,2, \ldots K$ et $E_0^N = POP - R_N$

par exemple = $E_4^5 = + 5 * [-R_5 + 4R_4 - 6R_3 + 4R_2 - R_1]$

$$E_4^5 = 5 * [-500 + 1840 - 2460 + 1380 - 250] = 50$$

Les deux tableaux suivants nous décrivent le processus utilisé pour obtenir les conversions hypergéométriques directes et inverses.

Conversion de la distribution des occasions de voir en la couverture nette cumulée

Individus exposés à		Coefficients hypergéométriques directs				
		R_1	R_2	R_3	R_4	R_5
0 sur 5	E_0^5 = 500	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
1 sur 5	E_1^5 = 200	1/5 (40)	2/5 (80)	3/5(120)	4/5 (160)	1 (200)
2 sur 5	E_2^5 = 100	2/5 (40)	7/10 (70)	9/10(90)	1 (100)	1 (100)
3 sur 5	E_3^5 = 50	3/5 (30)	9/10 (45)	1 (50)	1 (50)	1 (50)
4 sur 5	E_4^5 = 50	4/5 (40)	1 (50)	1 (50)	1 (50)	1 (50)
5 sur 5	E_5^5 = 100	1(100)	1 (100)	1 (100)	1 (100)	1 (100)
Couverture nette cumulée	=	R_1 = 250	R_2 = 345	R_3 = 410	R_4 = 460	R_5 = 500

(Les chiffres entre parenthèses représentent le nombre d'individus multiplié par les coefficients hypergéométriques directs).

Conversion de la couverture nette cumulée en distribution des occasions de voir

Couverture nette cumulée		Coefficients hypergéométriques inverses				
Parutions	Couverture nette	E_5^5	E_4^5	E_3^5	E_2^5	E_1^5
1	R_1 = 250	+5 (+1250)	-5 (-1250)			
2	R_2 = 345	-10 (-3450)	+20 (+6900)	-10 (-3450)		
3	R_3 = 410	+10 (+4100)	-30(-12300)	+30(+12300)	-10 (-4100)	
4	R_4 = 460	-5 (-2300)	+20 (+9200)	-30(-13800)	+20 (+9200)	-5(-2300)
5	R_5 = 500	+1 (+500)	-5 (-2500)	+10 (+5000)	-10 (-5000)	+5(+2500)
Distribution des occasions de voir		E_5^5 = 100	E_4^5 = 50	E_3^5 = 50	E_2^5 = 100	E_1^5 = 200

(Les chiffres entre parenthèses représentent le nombre d'individus multipliés par les coefficients hypergéométriques inverses).

Nous avons ainsi démontré que, dans l'hypothèse de parutions équiprobables, la couverture nette cumulée peut être convertie en distribution des occasions de voir.

En résumé, quand les probabilités individuelles d'exposition à un support donné sont indépendantes de la parution ainsi que des expositions passées, on peut convertir la distribution des occasions de voir en couverture nette cumulée et inversément.

Cependant, en pratique, ces deux hypothèses n'ont jamais été dûment testées.

Suivant ces deux hypothèses, l'audience au numéro devrait être constante, à condition qu'aucun changement n'intervienne dans la population. Mais il a été montré que l'audience au numéro est saisonnière et que, de plus, elle est sensible aux évènements d'importance nationale. En conséquence, l'hypothèse de stationnarité des probabilités individuelles d'exposition est sujette à question.

Néanmoins, lorsque l'on possède les données portant simultanément sur l'audience cumulée et la distribution des occasions de voir, on constate que la conversion directe des secondes dans les premières reproduit fidèlement l'audience cumulée.

Cela n'est pas nécessairement vrai de la conversion inverse qui va de l'audience cumulée à la distribution des occasions de voir. De faibles erreurs, dans la mesure des audiences cumulées sont amplifiées par les coefficients hypergéométriques inverses. Cela peut conduire à l'apparition de fréquences négatives.

Par exemple, si l'on considère les diverses tranches d'âge qui figurent dans l'enquête faite en 1958 par Politz pour le magazine Life, on n'aura pas de difficulté à convertir les données correspondant à l'ensemble des adultes (10 ans et plus). Par contre, si l'on convertit séparément chaque classe d'âge, les résultats seront incorrects pour les tranches d'âge 30-39, 40-49, 50-64 et 64 et plus, ainsi que le montre la table suivante.

Distribution des occasions de voir par classe d'âge
(Calculée à partir des audiences cumulées par application de la conversion hypergéométrique inverse)

Fréquence d'exposition	Population totale	de 10 à 19 ans	de 20 à 29 ans	de 30 à 39 ans*	de 40 à 49 ans*	de 50 à 64 ans*	65 ans et plus
0 sur 6	64,800	10,800	9,250	11,950	10,150	13,250	9,400
1 sur 6	20,100	4,500	3,300	4,200	3,300	3,300	1,500
2 sur 6	12,750	3,000	2,250	3,000	2,250	1,500	750
3 sur 6	10,000	3,000	2,000	1,000	1,000	3,000	-0
4 sur 6	6,000	750	750	3,750	750	- 1,500	1,500
5 sur 6	6,900	900	1,500	- 300	2,400	3,300	- 900
6 sur 6	5,800	1,350	1,100	1,900	400	200	850
Total	126,350	24,300	20,150	25,150	25,250	23,050	13,100

On remarque que, en raison des erreurs d'échantillonnage, chaque classe négative correspond à une exagération dans la classe d'âge suivante et précédente.

Ayant ainsi examiné la relation qui unit l'audience cumulée à la distribution des occasions de voir, nous allons étudier comment ces deux phénomènes sont reliés à leur cause commune : les différents degrés de fidélité présentés par les individus composant l'audience du support.

2 - Les relations liant la fidélité au support à l'accumulation et à la distribution des occasions de voir.

L'audience fidèle d'un support sera définie comme étant le nombre d'individus qui ont été exposés à exactement N parutions sur N. Ce concept n'est pas nouveau puisqu'il a été effleuré par Adam (3) et Broadbent sous le nom de régularité de lecture (4) et par Hofmans sous le nom de lecteurs ultraloyaux (5).

Toutefois, aucun de ces auteurs n'a donné un traitement mathématique complet de la relation entre la fidélité, l'accumulation et la fréquence d'exposition, sous l'hypothèse de la constance de l'audience au numéro.

21 - De la couverture nette cumulée à la distribution des audiences fidèles.

Supposons que l'audience cumulée générée par cinq insertions dans un support donné soit la suivante :

Parutions	1	2	3	4	5
Couverture nette cumulée	250	345	410	460	500

Supposons en outre que les parutions sont équiprobables, nous pouvons, à l'aide de la formule(2), convertir la couverture nette en la distribution des occasions de voir. Toutefois, nous nous occupons ici du nombre d'individus ayant été exposés à t parutions parmi t. En conséquence, la formule (2) se simplifie et nous obtenons :

$$(3) \qquad E_t^t = \sum_{K=1}^{t} (-1)^{K+1} \binom{t}{K} R_K \quad \text{avec } 1 \leq t \leq N \quad \text{et } K = 1, 2, \ldots, t.$$

Cette formule montre que la distribution de l'audience fidèle peut être dérivée de la couverture nette cumulée. Tout ce qui est nécessaire, c'est de prémultiplier le vecteur de l'audience cumulée R par la matrice B des coefficients binomiaux en affectant les colonnes paires du signe moins.

$$\underset{(N,N)}{B} * \underset{(N,1)}{R} = \underset{(N,1)}{L}$$

Appliquons ce résultat aux données de notre exemple

$$\begin{pmatrix} 1 & & & & \\ 2 & -1 & & & \\ 3 & -3 & 1 & & \\ 4 & -6 & 4 & -1 & \\ 5 & -10 & 10 & -5 & 1 \end{pmatrix} * \begin{pmatrix} 250 \\ 345 \\ 410 \\ 460 \\ 500 \end{pmatrix} = \begin{pmatrix} 250 \\ 155 \\ 125 \\ 110 \\ 100 \end{pmatrix}$$

22 - De la distribution des audiences fidèles à la couverture nette cumulée.

Supposons que la distribution des audiences fidèles à cinq parutions soit la suivante :

Parutions	1	2	3	4	5
Audiences fidèles	250	155	125	110	100

Nous cherchons à calculer l'audience cumulée à partir de ce que nous savons sur la distribution des audiences fidèles. Ce problème est l'inverse du problème précédemment résolu. Utilisant la formule (3), nous pouvons écrire cinq équations où les cinq audiences cumulées inconnues sont une fonction linéaire des cinq audiences fidèles. Nous obtenons :

$$E_1^1 = R_1 = 250$$

$$E_2^2 = 2R_1 - R_2 = 155$$

$$E_3^3 = 3R_1 - 3R_2 + R_3 = 125$$

$$E_4^4 = 4R_1 - 6R_2 + 4R_3 - R_4 = 110$$

$$R_5^5 = 5R_1 - 10R_2 + 10R_3 - 5R_4 + R_5 = 100$$

Nous résolvons ce système de cinq équations à cinq inconnues par l'inversion de la matrice des coefficients binomiaux B

$$\underset{(N,1)}{R} = \underset{(N,N)}{B^{-1}} * \underset{(N,1)}{L}$$

En raison de la nature particulière de ses coefficients, la matrice triangulaire B est égale à sa propre inverse. Nous obtenons ainsi le vecteur R, en effectuant la multiplication matricielle suivante :

$$\begin{pmatrix} 1 & & & & \\ 2 & -1 & & & \\ 3 & -3 & 1 & & \\ 4 & -6 & 4 & -1 & \\ 5 & -10 & 10 & -5 & 1 \end{pmatrix} * \begin{pmatrix} 250 \\ 155 \\ 125 \\ 110 \\ 100 \end{pmatrix} = \begin{pmatrix} 250 \\ 345 \\ 410 \\ 460 \\ 500 \end{pmatrix}$$

En conséquence, nous pouvons passer de la distribution des audiences loyales à l'audience cumulée à l'aide de la formule suivante :

$$(4) \qquad R_t = \sum_{K=1}^{t} (-1)^{K+1} \binom{t}{K} E_K^K$$

avec $1 \leq t \leq N$

et $K = 1, 2, \ldots, t$

Les deux tables suivantes décrivent, point par point, les conversions binomales directes et inverses.

Conversion de la couverture nette cumulée en la distribution des audiences fidèles

Couverture nette cumulée		Coefficients binomiaux directs				
Parutions	Couverture	E_1^1	E_2^2	E_3^3	E_4^4	E_5^5
1	$R_1 = 250$	1(250)	2(500)	3(750)	4(1000)	5(1250)
2	$R_2 = 345$		-1(-345)	-3(-1035)	-6(-2070)	-10(-3450)
3	$R_3 = 410$			1(410)	4(1640)	10(4100)
4	$R_4 = 460$				-1(-460)	-5(-2300)
5	$R_5 = 500$					1(500)
Audiences fidèles =		$E_1^1 = 250$	$E_2^2 = 155$	$E_3^3 = 125$	$E_4^4 = 110$	$E_5^5 = 100$

(Les chiffres entre parenthèses représentent le nombre d'indivus multiplié par les coefficients binomiaux directs).

Conversion de la distribution des audiences fidèles en la couverture nette cumulée

Audiences fidèles		Coefficients binomiaux inverses				
Individus exposés à		R_1	R_2	R_3	R_4	R_5
1 sur 1	$E_1^1 = 250$	+ 1(250)	+ 2(500)	+3(750)	+4 (+1000)	+5(1250)
2 sur 2	$E_2^2 = 155$		-1(155)	-3(-465)	-6(-930)	-10(-1550)
3 sur 3	$E_3^3 = 125$			+1 (125)	+4(+500)	+10(+1250)
4 sur 4	$E_4^4 = 110$				-1(-110)	-5(-550)
5 sur 5	$E_5^5 = 100$					+1(+100)
Audiences cumulées		$R_1 = 250$	$R_2 = 345$	$R_3 = 410$	$R_4 = 460$	$R_5 = 500$

(Les chiffres entre parenthèses représentent le nombre d'individus multipliés par les coefficients binomiaux inverses).

23 - De la distribution des occasions de voir à la distribution des audiences fidèles.

Supposons que la distribution des occasions de voir soit la suivante :

Fréquence d'exposition	0	1	2	3	4	5
Nombre d'individus	500	200	100	50	50	100

Nous avons déjà démontré que l'on peut convertir la distribution des occasions de voir en la couverture nette cumulée, puis celle-ci en la distribution des audiences fidèles. En conséquence, on doit pouvoir établir une relation directe entre la distribution des occasions de voir et la distribution des audiences fidèles. Afin d'expliciter cette relation, il suffit de remplacer R_K

dans la formule (3) par sa valeur calculée à l'aide de la formule (1). Nous obtenons :

$$E_t^t = \Sigma_{K=1}^{T} (-1)^{K+1} \binom{K}{t} \left[\Sigma_{t=1}^{N} E_t^N \left(1 - \binom{N-K}{t} \Big/ \binom{N}{t}\right) \right]$$

En simplifiant, il vient :

$$(5) \quad E_t^t = \frac{1}{\binom{N}{t}} \Sigma_{K=t}^{N} \binom{K}{t} E_K^N$$

avec $1 \leq t \leq N$
et $K = t, t+1, \ldots, N$

Puisque t doit être supérieur ou égal à 1, il n'est pas nécessaire de connaître le nombre d'individus non exposés, E_0^N.

En appliquant la formule (5), nous obtenons :

$$\begin{bmatrix} 1/5 & & & & \\ & 1/10 & & & \\ & & 1/10 & & \\ & & & 1/5 & \\ & & & & 1 \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ & 1 & 3 & 6 & 10 \\ & & 1 & 4 & 10 \\ & & & 1 & 5 \\ & & & & 1 \end{bmatrix} * \begin{bmatrix} 200 \\ 100 \\ 50 \\ 50 \\ 100 \end{bmatrix} = \begin{bmatrix} 250 \\ 155 \\ 125 \\ 110 \\ 100 \end{bmatrix}$$

24 - <u>De la distribution des audiences fidèles à la distribution des occasions de voir</u>.

Ce problème est l'inverse du problème précédent.

Supposons que la distribution des audiences fidèles à cinq parutions soit la suivante :

Parutions	1	2	3	4	5
Audiences fidèles	250	155	125	110	100

A l'aide de la formule (5), nous pouvons écrire cinq équations où les cinq fréquences d'exposition sont une fonction linéaire des cinq audiences fidèles.

$$
\begin{array}{lllllllll}
E_1^1 & = & 1/5E_1^5 & + & 2/5E_2^5 & + & 3/5E_3^5 & + & 4/5E_4^5 & + & E_5^5 & = & 250 \\
E_2^2 & = & & & 1/10E_2^5 & + & 3/10E_3^5 & + & 6/10E_4^5 & + & E_5^5 & = & 155 \\
E_3^3 & = & & & & & 1/10E_3^5 & + & 4/10E_4^5 & + & E_3^5 & = & 125 \\
E_4^4 & = & & & & & & & 1/5E_4^5 & + & E_5^5 & = & 110 \\
E_5^5 & = & & & & & & & & & E_5^5 & = & 100
\end{array}
$$

Nous résolvons ce système par l'inversion de la matrice des coefficients. Utilisant la formule (5), la matrice des coefficients peut s'écrire comme suit :

$$
AB = \begin{bmatrix} 1/\binom{N}{1} & & & & \\ & 1/\binom{N}{2} & & & \\ & & \ddots & & \\ & & & 1/\binom{N}{K} & \\ & & & & \ddots \\ & & & & & 1/\binom{N}{N} \end{bmatrix} * \begin{bmatrix} \binom{1}{1} & \binom{2}{1} \cdots\cdots \binom{K}{1} \cdots\cdots \binom{N}{1} \\ & \binom{2}{2} \cdots\cdots \binom{K}{2} \cdots\cdots \binom{N}{2} \\ & \ddots \quad \binom{K}{K} \cdots\cdots \binom{N}{K} \\ & \ddots \quad \ddots \quad \binom{N}{N} \end{bmatrix}
$$

L'inverse du produit de deux matrices est égal au produit commuté de leurs inverses :

$$
(AB)^{-1} = B^{-1} * A^{-1}
$$

En conséquence, l'inverse de la matrice des coefficients est donné par :

$$
\begin{bmatrix} \binom{1}{1} \cdots\cdots \binom{K}{1} \cdots\cdots \binom{N}{1} \\ \ddots \quad \vdots \quad \vdots \\ \binom{K}{K} \cdots\cdots \binom{N}{K} \\ \ddots \quad \vdots \\ \binom{N}{N} \end{bmatrix} * \begin{bmatrix} \binom{N}{1} & & \\ & \binom{N}{K} & \\ & & \ddots \\ & & & \binom{N}{N} \end{bmatrix} = (-1)^{K+t} \begin{bmatrix} \binom{N}{1}\binom{1}{1} \cdots\cdots \binom{N}{K}\binom{K}{1} \cdots\cdots \binom{N}{N}\binom{N}{1} \\ \ddots \quad \vdots \quad \vdots \\ \binom{N}{K}\binom{K}{K} \cdots\cdots \binom{N}{N}\binom{N}{K} \\ \ddots \quad \vdots \\ \binom{N}{N}\binom{N}{N} \end{bmatrix}
$$

Ainsi, nous pouvons passer des audiences fidèles à la distribution des occasions de voir par la formule suivante :

$$(6) \quad E_K^N = \Sigma_{t=K}^{N} (-1)^{K+t} \binom{N}{t}\binom{t}{K} E_t^t \qquad \text{avec} \quad 1 \leq K \leq N \quad \text{et} \quad t = K, K+1, \ldots, N$$

Le nombre d'individus non exposés s'obtient aisément en calculant :

$$E_0^N = POP - \Sigma_{K=1}^{N} E_K^N$$

En appliquant la formule (6) à la distribution des audiences fidèles, nous obtenons :

$$\begin{pmatrix} 1 & -2 & 3 & -4 & 5 \\ & 1 & -3 & 6 & -10 \\ & & 1 & -4 & 10 \\ & & & 1 & -5 \\ & & & & 1 \end{pmatrix} * \begin{pmatrix} 5 & & & & \\ & 10 & & & \\ & & 10 & & \\ & & & 5 & \\ & & & & 1 \end{pmatrix} * \begin{pmatrix} 250 \\ 155 \\ 125 \\ 110 \\ 100 \end{pmatrix} = \begin{pmatrix} 200 \\ 100 \\ 50 \\ 50 \\ 100 \end{pmatrix}.$$

Avec : $E_0^5 = 1000 - (200 + 100 + 50 + 50 + 100) = 500$

Ceci termine notre étude des relations liant la fidélité de l'audience à l'accumulation et à la distribution des occasions de voir. Nous avons démontré que la connaissance de l'une quelconque de ces données permet de calculer les deux autres, sous l'hypothèse de la constance de l'audience au numéro.

Les deux tables suivantes décrivent, point par point, la conversion de la distribution des occasions de voir en la distribution des audiences fidèles et inversément.

Conversion de la distribution des occasions de voir en la distribution des audiences fidèles

Occasions de voir / Individus exposés à		E_1^1	E_2^2	Coefficients de conversion E_3^3	E_4^4	E_5^5
1 sur 5	$E_1^5 = 200$	1/5(40)				
2 sur 5	$E_2^5 = 100$	2/5(40)	1/10(10)			
3 sur 5	$E_3^5 = 50$	3/5(30)	3/10(15)	1/10(5)		
4 sur 5	$E_4^5 = 50$	4/5(40)	6/10(30)	4/10(20)	1/5(10)	
5 sur 5	$E_5^5 = 100$	5/5(100)	10/10(100)	10/10(100)	5/5(100)	+1(100)
Audiences fidèles=		$E_1^1 = 250$	$E_2^2 = 155$	$E_3^3 = 125$	$E_4^4 = 110$	$E_5^5 = 100$

(Les chiffres entre parenthèses représentent le nombre d'individus multiplié par les coefficients de conversion).

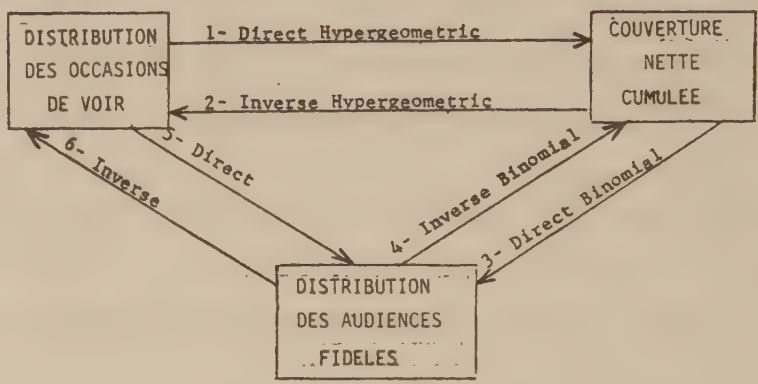
Conversion de la distribution des audiences fidèles en la distribution des occasions de voir

Audiences fidèles / Individus exposés à		E_1^5	E_2^5	E_3^5	E_4^5	E_5^5
1 sur 1	$E_1^1 = 250$	5(1250)				
2 sur 2	$E_2^2 = 155$	-20(-3100)	10(-1550)			
3 sur 3	$E_3^3 = 125$	30(3750)	-30(-3750)	+ 10(1250)		
4 sur 4	$E_4^4 = 110$	-20(-2200)	30(3300)	-20(-2200)	5(550)	
5 sur 5	$E_5^5 = 100$	5(500)	-10(+1000)	+10(+1000)	-5(-500)	1(100)
Occasions de voir =		$E_1^5 = 200$	$E_2^5 = 100$	$E_3^5 = 50$	$E_4^5 = 50$	$E_5^5 = 100$

(Les chiffres entre parenthèses représentent le nombre d'individus

Les relations entre les trois concepts de couverture nette cumulée, de distribution des occasions de voir et de distribution des audiences fidèles sont résumées dans la table suivante :

Relations entre la couverture nette cumulée, la distribution des occasions de voir et la distribution des audiences fidèles



1 - $R_t = \sum_{K=1}^{N} \left[1-\binom{N-t}{K}/\binom{N}{K}\right] E_K^N \qquad 1 \leq t \leq N \;,\; 1 \leq K \leq N$

2 - $E_K^N = \binom{N}{K} \sum_{t=0}^{K} (-1)^{K+t+1} \binom{K}{t} R_{N-t} \qquad 0 \leq t \leq K \;,\; 1 \leq K \leq N$

3 - $E_t^t = \sum_{K=1}^{t} (-1)^{K+1} \binom{t}{K} R_K \qquad 1 \leq t \leq N \;,\; 1 \leq K \leq t$

4 - $R_t = \sum_{K=1}^{t} (-1)^{K+1} \binom{t}{K} E_K^K \qquad 1 \leq t \leq N \;,\; 1 \leq K \leq t$

5 - $E_t^t = \left[1/\binom{N}{t}\right] \sum_{K=t}^{N} \binom{K}{t} E_K^N \qquad 1 \leq t \leq N \;,\; t \leq K \leq N$

6 - $E_k^N = \sum_{t=K}^{N} (-1)^{K+t} \binom{N}{t} \binom{t}{K} E_t^t \qquad K \leq t \leq N \;,\; 1 \leq K \leq N$

Les conversions, qu'elles soient directes ou inverses, nécessitent la collecte préalable de données, soit sur la distribution des occasions de voir, soit sur la couverture nette cumulée ou soit sur la distribution des audiences fidèles. Dès que l'une quelconque de ces données a été recueillie, pour le nombre de parutions désirées, il est possible de calculer les deux autres séries de données.

Toutefois, ces données sont coûteuses et leur obtention met en cause toute une série de difficultés qui ont été examinées dans le chapitre 4. Par exemple, les enquêtes de Politz, en 1953 et en 1958, pour le Magazine Life, ont nécessité six vagues d'interviews, chacune d'elle conduite par la méthode de la lecture au numéro, afin de déterminer la distribution des occasions de voir six parutions. Il est évident que la mesure directe d'un nombre plus élevé de parutions est matériellement impossible.

En conséquence, le besoin s'est fait sentir de développer des modèles d'accumulation d'audience capable d'extrapoler la couverture nette pour un nombre quelconque de parutions sur la base de l'audience des deux premières parutions.

Nous allons donc examiner maintenant les divers modèles d'accumulation d'audience qui ont été proposés. Nous examinerons leurs performances empiriques et nous discuterons la validité de leurs hypothèses, ce qui nous permettra de construire de nouveaux modèles aux performances supérieures.

3 - Les modèles d'accumulation d'audience.

Ces modèles peuvent être classés en trois catégories.

Les modèles de type Ad-Hoc, qui se contentent d'ajuster les données disponibles en vue d'extrapoler, les modèles stochastiques qui s'efforcent de réduire les comportements d'exposition à diverses lois de probabilité en recourant à plusieurs hypothèses simplificatrices, et les modèles de simulation qui s'appliquent à reproduire la réalité en affectant à chaque individu une probabilité personnelle d'exposition.

31. Les modèles Ad-Hoc.

Si l'on trace la courbe de la couverture nette cumulée en fonction du

nombre de parutions, on observe que l'audience s'accroît mais à un taux décroissant. En conséquence, toute fonction mathématique du nombre de parutions qui possède cette propriété peut être ajustée aux données disponibles. Toutefois, les extrapolations basées sur un simple ajustement peuvent conduire à des résultats absurdes comme on le verra bientôt.

311. L'accumulation logarithmique.

La fonction logarithmique peut être ajustée selon la méthode des moindres carrés, aux données empiriques. Les deux tables suivantes montrent que les données empiriques se conforment remarquablement bien au modèle logarithmique suivant :

$$(7) \qquad R_N = a + b \log (N) \qquad \text{avec} \qquad N \geqslant 1$$

La première colonne extraite des enquêtes d'audience réalisées aux USA par Politz en 1953 et 1958 (1), présente le pourcentage cumulé de foyers couverts par six parutions pour divers magazines. La deuxième colonne présente les estimations obtenues par l'ajustement du modèle logarithmique.

On observe que les dix ajustements sont presque parfaits. En première analyse, le modèle logarithmique apparaît donc comme satisfaisant. Toutefois, ceci n'est qu'une apparence et voici pourquoi .

En premier lieu, l'audience moyenne du numéro, telle qu'elle résulte de l'application du modèle, ne coïncide pas nécessairement avec l'audience moyenne observée. Il en est ainsi puisque la méthode des moindres carrés minimise la somme du carré des différences et non pas une différence particulière. Ceci est une faiblesse du modèle puisque les seules données couramment disponibles pour extrapoler l'audience cumulée sont l'audience moyenne au numéro et l'audience cumulée de deux parutions. Ce problème pourrait être résolu en contraignant la régression de façon à ce que l'audience moyenne estimée soit égale à l'audience moyenne observée, mais le principal problème de la cohérence logique des extrapolations du modèle ne serait pas résolu pour autant, comme on va le voir.

Supposons que les seules données disponibles soient R_1 et R_2. En appliquant l'équation (7), nous obtenons :

$$\begin{cases} R_1 = a \\ R_2 = a + b \log (2) \end{cases}$$

Couverture nette cumulée de cinq magazines
(Enquête Politz - % de foyers exposés)

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
1	22,1	22,0	11,8	11,3	15,1	14,3	9,6	9,2	19,2	19,5
2	32,4	32,6	17,8	18,2	23,9	24,7	14,4	14,8	25,9	25,7
3	39,1	38,8	21,8	22,3	30,0	30,7	17,7	18,0	29,6	29,3
4	44,0	43,3	25,0	25,1	34,8	35,0	20,2	20,3	32,0	31,9
5	47,7	46,7	27,5	27,4	38,6	38,4	22,2	22,1	33,8	33,9
6	50,6	49,5	29,6	29,2	41,8	41,1	23,9	23,6	35,3	35,5
$\hat{a}$	0,220		0,113		0,143		0,092		0,195	
$\hat{b}$	0,353		0,229		0,344		0,184		0,207	
R^2	0,989		0,996		0,995		0,996		0,998	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
1	31,2	31,3	19,4	19,0	23,8	23,5	19,4	19,3	36,0	36,2
2	42,6	42,6	27,4	27,8	34,4	34,8	27,3	27,5	46,5	46,3
3	49,4	49,3	32,6	33,3	41,2	41,3	32,2	38,2	52,5	52,3
4	54,1	54,0	36,5	36,6	46,0	46,0	35,6	35,6	56,6	56,5
5	57,7	57,7	39,6	39,5	49,7	49,6	38,3	38,2	59,7	59,8
6	60,5	60,7	42,1	41,8	52,7	52,6	40,4	40,4	62,2	62,4
$\hat{a}$	0,313		0,190		0,235		0,193		0,362	
$\hat{b}$	0,378		0,293		0,373		0,271		0,337	
R^2	0,999		0,998		0,999		0,999		0,999	

En conséquence, les deux paramètres a et b de la régression sont donnés dans ce cas par :

$$(8) \qquad \begin{cases} \hat{a} = R_1 \\ \hat{b} = (R_2 - R_1) / \log (2) \end{cases}$$

La couverture nette cumulée, extrapolée à N parutions est alors égale à :

$$\hat{R}_N = R_1 + (R_2 - R_1) \log (N) / \log (2)$$

Les deux tables suivantes comparent la couverture nette observée avec la couverture nette extrapolée pour 3, 4, 5 et 6 parutions.

On observe que les paramètres estimés sur la base des deux premières parutions ne diffèrent que très peu des paramètres estimés sur la base des six parutions.

La validité prédictive du modèle logarithmique apparaît comme satisfaisante lorsque l'on ne considère que les six premières parutions. Toutefois, le modèle logarithmique n'admet pas d' asymptote supérieure. Il n'y a donc rien qui s'oppose à l'obtention d'une couverture nette supérieure à cent pour cent, ce qui, bien entendu, est logiquement impossible. Cette incohérence logique des extrapolations du modèle peut être démontrée en utilisant par exemple l'équation du magazine Life pour l'année 1958 pour calculer le nombre de parutions à partir duquel la couverture nette sera supérieure à 100 pour 100.

$$0{,}313 + 0{,}378 \log (N) \geqslant 1$$
$$\log (N) \geqslant (1 - 0{,}313) / 0{,}378$$
$$N \geqslant 66$$

On pourrait argumenter du fait qu'aucun plan de campagne ne prévoit un tel nombre d'insertions dans un seul support publicitaire. A cela nous répondrions qu'un modèle satisfaisant ne peut pas se contenter de reproduire avec succès les observations disponibles. Il doit aussi satisfaire aux contraintes de cohérence logique propre au phénomène qu'il se propose d'expliquer. C'est ainsi que la couverture nette doit être nulle en l'absence de toute parution. Cependant,

Extrapolation logarithmique de la couverture nette cumulée sur la base de l'audience des deux premières parutions

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
3	39,1	38,4	21,8	21,3	30.0	29,0	17,7	17,2	29,6	29,8
4	44,0	42,7	25,0	23,8	34,8	32,7	20,2	19,2	32,0	32,6
5	47,7	46,0	27,5	25,7	38,6	35,5	22,2	20,7	33,8	34,8
6	50,6	48,7	29,6	27,3	41,8	37,8	23,9	22,0	35,3	36,5
$\hat{a}$	0,221		0,118		0,151		0,96		0,192	
$\hat{b}$	0,342		0,199		0,292		0,159		0,223	
Ecart type	0,136		0,159		0,275		0,132		0,0084	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		READERS DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
3	49,4	49,3	32,6	32,1	41,2	40,6	32,2	31,9	52,5	52,6
4	54,1	54,0	36,5	35,4	46,0	45,0	35,6	35,2	56,6	57,0
5	57,7	57,7	39,6	38,0	49,7	48,4	38,3	37,7	59,7	60,4
6	60,5	60,7	42,1	40,1	52,7	51,2	40,4	39,8	62,2	63,1
$\hat{a}$	0,312		0,194		0,238		0,194		0,350	
$\hat{b}$	0,379		0,266		0,352		0,262		0,349	
Ecart type	0,0012		0,0143		0,0115		0,0047		0,0062	

le modèle logarithmique indique une audience négative (e. g. moins l'infini) ! En fait, ce modèle est un bon exemple du dilemne classique opposant prédiction et explication. Finalement, le modèle logarithmique ne nous permet pas d'estimer directement la distribution des occasions de voir.

312. L'accumulation exponentielle.

Parmi les fonctions mathématiques présentant un taux de croissance décroissant, la fonction exponentielle suivante est particulièrement simple.

$$(9) \qquad R_N = RMAX - ab^N \qquad \text{avec} \qquad b < 1$$

Ce modèle possède l'avantage, par rapport au modèle logarithmique, d'admettre une assymptote supérieure RMAX. Toutefois, il présente l'inconvénient de dépendre de trois paramètres, a, b, et RMAX et donc de nécessiter la connaissance de la couverture nette cumulée des trois premières parutions. Si nous supposons que R_1, R_2, R_3 sont connus, les trois paramètres sont donnés par :

$$(10) \quad \begin{cases} \hat{a} = (RMAX - R_1)^2 \ / \ (RMAX - R_2) \\ \hat{b} = (RMAX - R_2) \ / \ (RMAX - R_1) \\ RMAX = (R_2^2 - R_1R_3) \ / \ (2R_2 - R_1 - R_3) = R_1 + ab \end{cases}$$

Il est possible de réécrire le modèle en faisant apparaître de façon explicite ses deux bornes, R_1, l'audience au numéro et, RMAX, l'audience maximum.

$$R_N = R_1 - (RMAX - R_1)(1 - e^{c(N-1)}) \qquad \text{avec} \qquad c < 1$$

Dans ce cas, les deux paramètres RMAX et c sont estimés comme suit

$$\begin{cases} RMAX = R_1 + (R_2 - R_1)(1 - e^c) \\ \hat{c} = Ln \left[(R_1 - R_3) \ / \ (R_1 - R_2) - 1 \right] \end{cases}$$

Les deux tables suivantes comparent la couverture nette observée avec la couverture nette extrapolée pour 4,5 et 6 parutions.

Extrapolation exponentielle de la couverture nette cumulée

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		READERS DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
4	44,0	43,5	25,0	24,5	34,8	34,2	20,2	20,0	32,0	31,6
5	47,7	46,3	27,0	26,2	38,6	37,2	22,2	21,5	33,8	32,8
6	50,6	48,1	29,6	27,4	41,8	39,2	23,9	22,6	35,3	33,4
$\hat{a}$	0,453		0,270		0,414		0,223		0,271	
$\hat{b}$	0,651		0,667		0,693		0,687		0,552	
RMAX	51,6 %		29,8 %		43,8 %		24,9 %		34,2 % *	
Ecart type	0,0167		0,0148		0,0175		0,0085		0,0127	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		READERS DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
4	54,1	53,5	36,5	36,0	46,0	45,6	35,6	35,2	56,6	55,9
5	57,7	57,7	39,6	38,2	49,7	48,4	38,3	37,1	59,7	57,9
6	60,5	57,3	42,1	39,6	52,7	50,2	40,4	38,3	62,2	59,0
$\hat{a}$	0,474		0,352		0,461		0,335		0,429	
$\hat{b}$	0,596		0,650		0,642		0,620		0,571	
RMAX	59,4 % *		42,3 %		53,4 %		40,2 % *		60,5 % *	
Ecart type	0,215		0,0169		0,0168		0,0141		0,0215	

* L'étoile dénote des résultats en contradiction avec les données observées.

On observe à partir de ces tables que le modèle exponentiel sous estime systématiquement la couverture nette cumulée. L'écart type des prédictions du modèle est en moyenne de 0,0161 pour les dix magazines contre 0,0116 pour le modèle logarithmique. Il apparait donc que le modèle exponentiel, tout en nécessitant davantage d'informations que le modèle logarithmique, s'ajuste moins bien aux données. De plus, les estimations de la couverture nette maximum sont en contradiction avec les observations puisque, dans quatre cas sur dix, le maximum estimé est en fait inférieur à la couverture nette de six parutions !

En résumé, le modèle exponentiel, quoique théoriquement supérieur au modèle logarithmique puisqu'il admet une asymptote supérieure, se révèle inférieur en matière de validité prédictive puisque ses estimations sont en contradiction avec les observations. Ils ne permet pas non plus d'estimer directement la distribution des occasions de voir.

313. La formule d'Agostini.

Dans un article paru en 1962 (6), Agostini propose un modèle d'accumulation d'audience permettant d'extrapoler la couverture nette à partir de l'audience des deux premières parutions. Utilisant l'hypothèse de l'accumulation aléatoire (7), il montre, sur la base des enquêtes Politz de 1953 et 1958, que les audiences s'accumulent à un taux moindre que celui qui résulterait d'une exposition purement aléatoire.

Soit R_N, la couverture nette en pourcentage, atteinte après N parutions, et soit $\Delta_N = R_N - R_{N-1}$ l'accroissement de couverture dû à la Nième parution. Soit $UR_{N-1} = 1 - R_{N-1}$ le pourcentage de non couverture après N-1 parutions. Agostini observe que le rapport $y = \Delta_N / UR_{N-1}$, qui devrait être constant et égal à R_1 sous l'hypothèse d'accumulation aléatoire, en fait diminue lorsque N s'accroît. De plus, il montre qu'une équation du type $Y_N = a/N^b$, permet un bon ajustement de la relation empirique liant Y à N.

En conséquence, son modèle peut s'écrire :

$$Y_N = \frac{\Delta_N}{UR_{N-1}} = \frac{a}{N^b}$$

Pour $N = 1$ et $N = 2$, il vient

$$\begin{cases} R_1 = a \\ R_2 = R_1 + (1 - R_1) \dfrac{a}{2^b} \end{cases}$$

En conséquence, les deux paramètres du modèle d'Agostini sont donnés par :

$$\begin{cases} \hat{a} = R_1 \\ \hat{b} = \log \left[R_1 (1 - R_1) / (R_2 - R_1) \right] / \log (2) \end{cases}$$

La couverture nette cumulée à N parutions sera donc estimée par la formule suivante :

$$(12) \qquad R_N = 1 - \left[(1 - R_1) (1 - R_1/2^b) \ldots . (1 - R_1/N^b) \right]$$

Les deux tables suivantes montrent que cette équation permet un meilleur ajustement que celui obtenu par les modèles logarithmique et exponentiel. L'écart type moyen des erreurs de prédictions pour les dix magazines n'est plus que de 0,0061 contre 0,0116 pour le modèle logarithmique et 0,0161 pour le modèle exponentiel. De plus, le modèle d'Agostini ne peut jamais conduire à une couverture nette supérieure à cent pour cent. En effet, dans l'équation (12), la quantité entre crochet, qui est un produit de termes tous inférieurs à l'unité, tend vers zéro lorsque le nombre de parutions s'accroît.

Finalement, bien que la formule d'Agostini ne nous donne pas une théorie ou une explication du phénomène d'exposition, elle nous apprend, du moins, que l'accumulation des audiences s'écarte du modèle aléatoire d'une façon régulière et ordonnée. Cette observation ouvre la voie à de nouvelles recherches.

Extrapolation de la couverture nette cumulée par la formule d'Agostini

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
3	39,1	39,0	21,8	21,8	30,0	30,2	17,7	17,6	29,6	29,7
4	44,0	43,8	25,0	24,9	34,8	35,2	20,2	20,0	32,0	32,2
5	47,7	47,6	27.5	27,4	38,6	39,3	22,2	22,0	33,8	34,0
6	50,6	50,7	29,6	29,5	41,8	42,7	23,9	23,6	35,3	35,5
$\hat{a}$	0,221		0,118		0,151		0,096		0,0192	
$\hat{b}$	0,741		0,795		0,543		0,854		1,211	
Ecart type	0,0146		0,0011		0,0062		0,0021		0,0017	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
3	49,4	49,2	32,6	32.3	41,2	41,1	32,2	32,1	52,5	52,0
4	54,1	53,6	36,5	35.7	46,0	45,9	35,6	35,4	56,6	55,6
5	57,7	57,0	39,6	38.3	49,7	49,6	38,3	38,0	59,7	58,2
6	60,5	59,6	42,1	40.5	52,7	52,6	40,4	40,1	62,2	60,2
$\hat{a}$	0,312		0,194		0,238		0,194		0,360	
$\hat{b}$	0,913		0,967		0,775		0,985		1,134	
Ecart type	0,0064		0,0012		0,0014		0,0024		0,0137	

Agostini (8) a cherché à perfectionner sa formule d'accumulation en introduisant, de façon explicite, une limite supérieure M à la couverture nette. Ce nouveau paramètre permet de tenir compte d'une couverture nette maximum inférieure à cent pour cent et rend ainsi le modèle plus réaliste. En effet, ne serait-ce qu'en vertu de l'analphabétisme, on a tout lieu de croire qu'aucun support n'atteint la totalité d'une population, quel que soit le nombre de parutions utilisées.

Agostini émet l'hypothèse selon laquelle le rapport de la non couverture après N parutions à la non couverture après une parution suivrait une loi hyperbolique du type :

$$z = 1 / (K (N - 1) + 1).$$

Pour une parution, la non couverture est $M - R_1$, pour deux parutions, elle vaut $M - R_2$, en conséquence, le paramètre K est donné par :

$$K = (R_2 - R_1) / (M - R_2)$$

En portant cette valeur dans l'équation de z et en extrayant R_N, on obtient l'équation de la couverture nette cumulée après N parutions.

$$(13) \qquad R_N = M - \left[(M - R_2)(M - R_1)\right] \;/\; \left[(N - 1)(R_2 - R_1) + (M - R_2)\right]$$

Ce nouveau modèle peut être critiqué sur deux plans. En premier lieu l'audience n'est pas nulle lorsque N vaut zéro, c'est-à-dire lorsqu'il n'y a pas eu de parutions. En second lieu, la couverture nette maximum M est généralement inconnue. Toutefois, si l'on dispose de la couverture nette cumulée des trois premières parutions, ce qui est rarement le cas, M peut être estimé par :

$$(14) \qquad \hat{M} = (R_1R_2 + R_2R_3 - 2R_1R_3) \;/\; (2R_2 - R_1 - R_3)$$

Les deux tables suivantes décrivent l'ajustement obtenu en utilisant les équations (13) et (14).

Extrapolation de la couverture nette cumulée par la nouvelle formule d'Agostini

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
4	44,0	43,3	25,0	24,7	34,8	34,5	20,2	20,1	32,0	31,9
5	47,7	47,3	27,0	26,8	38,6	37,9	22,2	21,9	33,8	33,6
6	50,6	50,0	29,6	28,5	41,8	40,6	23,9	23,4	35,3	34,8
$\hat{M}$	70,7 %		41,8 %		63,7 %		35,5 %		42,4 %	
Ecart type	0,0043		0,0067		0,0082		0,0034		0,0032	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
4	54,1	53,9	36,5	36,3	46,0	45,9	35,6	35,5	56,6	56,4
5	57,7	57,1	39,6	39,0	49,7	49,4	38,3	38,0	59,7	59,1
6	60,5	59,5	42,1	41,0	52,7	52,1	40,4	39,8	62,2	61,1
$\hat{M}$	76,3 %		57,1 %		72,3 %		53,1 %		74,5 %	
Ecart type	0,0068		0,0073		0,0039		0,0039		0,0073	

L'écart type moyen des erreurs de prédiction pour les dix magazines est désormais de 0,0055 contre 0,0061 pour la première formule d'Agostini. De plus, les valeurs estimées de la couverture nette maximum ne sont pas en contradiction avec les données comme c'était le cas pour le modèle exponentiel. Néanmoins, on note que la nouvelle formule sous estime systématiquement la couverture nette ce qui n'était pas le cas pour la première formule d'Agostini. On peut donc conclure que lorsque l'on dispose d'une estimation fiable de la couverture nette maximum ou que l'on connaît la couverture nette des trois premières parutions, il est avantageux d'utiliser la seconde formule d'Agostini. Toutefois, aucune des deux formules ne permet d'estimer directement la distribution des occasions de voir.

314. La courbe d'accumulation d'Engelman.

Engelman, sur la base des travaux d'Agostini, entreprit la construction d'une courbe d'accumulation "idéale", valable pour tout magazine (9). Pour éliminer, les différences d'audience entre les magazines, il travaille sur la base de la fréquence moyenne d'exposition. Celle-ci est égale à l'audience brute de N parutions divisée par la couverture nette. Son évolution en fonction du nombre de parutions reflète uniquement le taux d'accumulation, indépendamment de la taille de l'audience.

Soit F_N la fréquence moyenne d'exposition à N parutions. Nous avons par définition :

$$F_N = R_1 * N / R_N \text{ , et en particulier, } \quad F_1 = 1$$

Le problème consiste maintenant à découvrir une fonction du nombre de parution susceptible de fournir un bon ajustement aux fréquences moyennes observées. La solution la plus simple consiste à choisir la fonction linéaire :

$$F_N = a + bN$$

Si nous supposons que seules F_1 et F_2 sont connues, les paramètres a et b sont donnés par :

$$\begin{cases} a = 1 - b \\ b = F_2 - 1 \end{cases}$$

Toutefois, cette approche échoue car les fréquences moyennes observées ne sont pas une fonction linéaire du nombre de parutions ainsi que le montre l'exemple suivant.

Couverture nette cumulée du Magazine LIFE - 1953

Parutions	1	2	3	4	5	6
Couverture nette cumulée	22,1	32,4	39,1	44,0	47,7	50,6
Fréquence moyenne observée	1	1,364	1,696	2,009	2,317	2,621
Fréquence moyenne estimée	1	1,364	1,728	2,092	2,456	2,820

Il apparaît clairement sur cet exemple que la courbe de fréquence moyenne s'accroît à un taux légèrement décroissant. Plutôt que d'ajuster aux données une fonction mathématique plus complexe, Engelman construit une courbe moyenne ou "idéale" basée sur la moyenne arithmétique des observations dont il dispose (10). Néanmoins, cet auteur n'indique pas l'ordre de grandeur de la variance existant entre les diverses courbes F_N empiriquement observées. En conséquence, il est difficile d'apprécier la validité de son approche (11). Le seul mérite que l'on puisse reconnaître au recours à une courbe standard tient à ce qu'elle permet de construire des courbes de couverture nette pour divers segments démographiques dont la somme est égale à la courbe de couverture nette de la population totale. Cette propriété d'homogénéité n'est pas vérifiée pour les autres méthodes d'extrapolation de la couverture nette.

315. La formule d'Hofmans.

Hofmans (5), dans un article important paru en 1966, présente un modèle d'accumulation d'audience, de type Ad Hoc, basé sur une modification du modèle de duplication d'audience proposé en 1961 par Agostini (12). En se basant sur l'analogie, déjà mentionnée, entre l'exposition à N supports et l'exposition à N parutions d'un même support, Hofmans suggère d'étendre la formule de duplication d'Agostini à la prédiction des audiences cumulées.

La formule de duplication d'Agostini sera étudiée en détail dans le chapître 6. Pour l'instant, il suffit de savoir que cette formule estime la

couverture nette de m supports en fonction de leur audience brute, des audiences dupliquées de chaque paire de supports et d'une constante k.

Nous adopterons la notation suivante :

A_i = audience du support i

$A = \Sigma_{i=1}^{m} A_i$ = audience brute des m supports.

$A_{ij} = A_i \Omega A_j$ = audience dupliquée de la paire de supports i,j.

$D = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}$ = somme des audiences dupliquées.

R_m = couverture nette de m supports.

Dans ces notations, la formule d'Agostini s'exprime par :

$$R_m = A^2 / (A + kD)$$

La transposition de la formule au cas de N parutions d'un même support s'obtient aisément car sous l'hypothèse de la constance de l'audience au numéro, l'audience brute est égale à N fois l'audience moyenne d'un numéro.

$$A = \Sigma_{i=1}^{N} A_i = N * R_1$$

Si nous admettons de plus que l'audience dupliquée de deux parutions est constante, quelles que soient les parutions considérées, nous obtenons :

$$D = N(N - 1)\, d/2 \quad \text{avec} \quad d = A_{ij} = A_{i'j'}$$

$$\forall_i, i' \in N \; ; \; \forall_j, j' \in N$$

$$N = 1, 2, \ldots, i, \ldots, j, \ldots, N$$

Nous pouvons calculer d, l'audience dupliquée de deux parutions quelconques, en fonction des audiences cumulées des deux premières parutions R_1 et R_2. En effet, l'audience cumulée des deux premières parutions est, par définition, égale à la somme des audiences des deux parutions moins l'audience dupliquée.

$$R_2 = 2 R_1 - d \qquad \text{d'où} : d = 2R_1 - R_2$$

En remplaçant, dans la formule d'Agostini, A et D par leur valeur, nous obtenons :

$$(15) \qquad R_N = (NR_1)^2 / [NR_1 + k\,N\,(N-1)\,(2R_1 - R_2) / 2]$$

Cette formule permet de calculer l'audience cumulée d'un nombre quelconque de parutions en fonction de R_1, R_2 et de la constante k. Toutefois, comme nous allons le voir, il apparait que la valeur de k, estimée à partir des données, n'est pas constante mais varie en fonction du nombre de parutions considérées.

Dans le cas de deux parutions, la formule (15) devient :

$$R_2 = (2\,R_1)^2 / [\,2\,R_1 + k\,(2\,R_1 - R_2)]$$

ce qui permet d'extraire la valeur de k

$$(16) \qquad k = 2\,R_1 / R_2$$

En utilisant cette estimation de la valeur de k, nous avons calculé pour nos données habituelles la couverture nette cumulée jusqu'à six parutions, sur la base de la formule (15). La table suivante compare ces estimations à la couverture nette observée. On peut remarquer que la formule (15) conduit à des couvertures nettes toujours inférieures à l'unité tant que $R_2 < 2\,R_1$. Cette condition est vérifiée pour tout support qui ne renouvelle pas entièrement son audience à chaque parution.

L'étude de cette table montre que l'ajustement aux données est inférieur à celui obtenu par le modèle logarithmique ou par le modèle d'Agostini : l'erreur type moyenne pour les 10 magazines est de 0,0215 contre 0,0116 pour le modèle logarithmique et 0,0061 pour le modèle d'Agostini. En outre, ce modèle sous-estime systématiquement la couverture nette réelle et cette sous-estimation s'aggrave lorsque le nombre de parutions s'accroit.

Mais, plus important encore, le paramètre k, supposé constant par hypothèse, varie avec le nombre de parutions considérées. Pour le montrer, considérons tout d'abord le cas de trois parutions. La formule (15) devient :

$$R_3 = (3\,R_1)^2 / (3\,R_1 + 3\,kd) = 3\,R_1^2 / (R_1 + kd)$$

Couverture nette observée et couverture nette estimée par le modèle de duplication constante

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
3	39,1	38,4	21,8	21,5	30,0	29,7	17,7	17,3	29,6	29,5
4	44,0	42,3	25,0	23,8	34,8	33,7	20,2	19,2	32,0	31,4
5	47,7	45,0	27,5	25,7	38,6	36,8	22,2	20,6	33,8	32,8
6	50,6	47,1	29,6	26,9	41,8	39,1	23,9	21,6	35,3	33,8
$\hat{d}$	11,8		5,8		6,3		4,8		12,5	
$\hat{k}$	1,36		1,33		1,26		1,33		1,48	
Ecart type	0,0239		0,0174		0,0172		0,0150		0,0095	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
3	49,4	48,5	32,6	31,8	41,2	40,4	32,2	31,6	52,5	51,5
4	54,1	52,1	36,5	34,5	46,0	44,3	35,6	34,3	56,6	54,4
5	57,7	54,6	39,6	36,4	49,7	46,9	38,3	36,1	59,7	56,4
6	60,5	56,3	42,1	37,8	52,7	49,0	40,4	37,5	62,2	57,7
$\hat{d}$	19,8		11,4		13,2		11,5		25,5	
$\hat{k}$	1,46		1,42		1,38		1,42		1,55	
Ecart type	0,0283		0,0289		0,0250		0,0196		0,0304	

En extrayant la valeur de k nous obtenons :

$$(17) \qquad k = [R_1 (3 R_1 - R_3)] \;/\; [R_3 (1 R_1 - R_2)]$$

Cette valeur est différente de celle obtenue en considérant deux parutions sauf si l'audience cumulée de trois parutions est exactement égale à :

$$R_3 = 3 R_1 R_2 \;/\; (4 R_1 - R_2)$$

Or, cette égalité n'est pas vérifiée par les données et l'hypothèse de la constance de k doit être rejetée (13). Nous pouvons généraliser ce résultat à un nombre quelconque de parutions en extrayant la valeur de k de la formule (15).

$$(18) \qquad k_N = [2 NR_1 (NR_1 - R_N)] \;/\; [N (N-1) (2 R_1 - R_2) R_N]$$

avec $N \geq 2$

Cette formule généralise les formules (16) et (17) à un nombre quelconque de parutions.

On peut donc affirmer que k n'est pas une constante, car cette hypothèse impose à la couverture nette cumulée de vérifier, quel que soit le nombre de parutions, l'égalité suivante :

$$(19) \qquad 2 R_1 / R_2 = 2 NR_1 (NR_1 - R_N) / N(N - 1) (2 R_1 - R_2) R_N$$

Bien entendu, il est absolument exclu que cette égalité puisse se vérifier en pratique pour toutes les valeurs de N et, en fait, elle n'est jamais vérifiée pour les données que nous avons examinées.

La variation de k en fonction du nombre de parutions étant acquise, l'étape suivante consiste à exprimer k_N par une fonction simple du nombre de

parutions. C'est ce qu'a réalisé Hofmans. En utilisant les données fournies par les enquêtes, Politz de 1953 et SOFRES de 1964, il conclut qu'une fonction de type puissance permet un bon ajustement des données. Il pose donc :

$$k_N = k_2 \, (N - 1)^{a-1} \qquad \text{avec} \qquad N \geq 2$$

Cependant, l'estimation du paramètre "a" nécessite la connaissance de l'audience cumulée des trois premiers numéros.

$$\hat{a} = \text{Log} \left[(3 R_1 - R_3) R_2 \; / \; (2 R_1 - R_2) R_3 \right] / \text{Log} (2)$$

Ceci constitue un handicap puisque les enquêtes d'audience se limitent le plus souvent à la mesure de l'audience cumulée de deux parutions.
En remplaçant k par k_N dans la formule (15), on obtient la formule d'Hofmans :

$$(20) \qquad R_N = (NR_1)^2 \; / \left[N R_1 + N (N - 1)^a (2 R_1 - R_2) R_1 / R_2 \right]$$

Les deux tables suivantes examinent les performances du modèle d'accumulation d'Hofmans sur la base de nos données habituelles.

L'examen de ces tables fait apparaître que le modèle de duplication variable d'Hofmans n'améliore que très peu la qualité de l'ajustement par rapport au modèle de duplication constante. L'erreur type moyenne de prédiction pour les dix magazines est de 0,0205 contre 0,0215 pour le modèle de duplication constante. Ce résultat est beaucoup moins bon que celui obtenu par le modèle d'Agostini dont l'erreur type moyenne s'élève seulement à 0,0061.

De plus, le modèle d'accumulation d'Hofmans surestime systématiquement la couverture nette cumulée réelle. Bien que plus satisfaisant sur le plan théorique, le modèle d'Hofmans s'avère donc inférieur, sur le plan empirique, au modèle logarithmique et au modèle d'Agostini. Finalement, le modèle d'Hofmans ne permet pas d'estimer directement la distribution des occasions de voir et nécessite la connaissance de l'audience cumulée à trois parutions alors que Simmons ne mesure que l'audience cumulée à deux parutions.

Nous en avons maintenant terminé avec les modèles d'accumulation de type Ad Hoc et nous entamons l'examen des modèles stochastiques d'accumulation d'audience.

Le modèle d'accumulation d'Hofmans

Comparaison de la couverture nette observée et de la couverture nette estimée

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
4	44,0	45,9	25,0	25,7	34,8	35,5	20,2	21,3	32,0	32,7
5	44,7	50,4	27,5	28,4	38,6	39,5	22,2	23,7	33,8	34,7
6	50,6	54,0	29,6	30,6	41,8	42,7	23,9	25,8	35,3	36,2
kd	0,161		0,077		0,079		0,064		0,185	
a	0,934		0,937		0,952		0,912		0,971	
Erreur type	0,0274		0,0086		0,0083		0,0153		0,0082	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
4	54,1	56,3	36,5	38,5	46,0	48,2	35,6	37,2	56,6	58,9
5	57,7	60,5	39,6	42,3	49,7	52,7	38,3	40,4	59,7	58,9
6	60,5	64,0	42,1	45,4	52,7	56,4	40,4	43,0	62,2	65,7
kd	0,289		0,162		0,182		0,163		0,395	
a	0,945		0,916		0,934		0,939		0,947	
Erreur type	0,0289		0,0270		0,0302		0,0212		0,0298	

32. Les modèles stochastiques.

Les comportements d'exposition peuvent être conçus comme un processus stochastique mettant en jeu deux entités statistiques, les individus et les parutions et un évènement aléatoire dichotomique, l'exposition ou la non-exposition. Nous allons analyser ici la construction d'énoncés probabilistes corrects, concernant l'occurrence d'un nombre spécifié d'expositions, étant donné un ensemble particulier d'hypothèses portant sur les individus et les parutions.

Les modèles stochastiques partagent, avec les modèles de simulation, l'hypothèse fondamentale que tous les individus de la population possèdent une probabilité personnelle d'exposition qui gouverne leur comportement de lecteur ou de spectateur. Dans le modèle le plus complet que l'on puisse imaginer, cette probabilité dépendrait à la fois de l'individu, du support, de la parution, des expositions passées et même des probabilités d'exposition des autres individus.

Cependant, un processus stochastique d'une telle complexité serait impossible à manier. Nous pourrons le simplifier grandement en recourant à l'hypothèse de stationnarité :-les évènements du passé ne modifient pas les mécanismes en action dans le modèle ; et à l'hypothèse d'homogénéité : -les individus se comportent de la même manière que la population totale.

Les hypothèses de stationnarité sont de deux sortes :

a) - L'hypothèse de stationnarité par rapport aux parutions rend la probabilité individuelle d'exposition indépendante des expositions passées ainsi que de la nature particulière de la chronologie et du contenu de la parution. En d'autres termes, la probabilité individuelle d'exposition à une parution est indépendante de l'historique des expositions précédentes comme du contenu éditorial de la parution.

b) - L'hypothèse de stationnarité par rapport aux individus rend la probabilité d'exposition de l'individu i indépendante de la probabilité d'exposition de l'individu j. En d'autres termes, cette hypothèse exclut la possibilité que l'exposition de l'individu i soit provoquée par une action de l'individu j ou procède d'une imitation du comportement de l'individu j.

La combinaison de ces deux hypothèses de stationnarité transforme les comportements de lecture et d'écoute en processus stochastique d'ordre zéro. La plupart des modèles stochastiques d'accumulation sont de ce type, cependant, nous

étudierons aussi un modèle d'accumulation Markovien d'ordre un.

En vue de procéder à une simplification ultérieure, certains modèles d'accumulation recourent à l'hypothèse de l'homogénéité des individus constituant la population cible. Ces modèles considèrent que toutes les probabilités individuelles d'exposition sont identiques. Parfois, l'existence de différences individuelles est reconnue, mais l'on considère toutefois que la variance autour de la probabilité moyenne d'exposition est négligeable, ce qui revient au même. Néanmoins, les modèles d'accumulation les plus satisfaisants reconnaissent explicitement l'hétérogénéité des individus et la modèlent soit par une segmentation des auditeurs, soit par la spécification d'une distribution des individus le long du continuum des probabilités d'exposition.

Il est donc possible de classer les divers modèles stochastiques d'accumulation, par ordre de complexité croissante, selon le nombre et la nature de leurs hypothèses quant aux probabilités d'exposition.

Classification des modèles stochastiques d'accumulation

		La population est : Homogène	Hétérogène
Les probabilités d'exposition sont :	Stationnaires	Binomial	Binomial segmenté Politz Beta-Binomial Négatif-Binomial Broadbent Morgensztern Béta-Binomial-Segmenté
	Non Stationnaires	Markov-Binomial	Simulation

En résumé, les modèles stochastiques d'accumulation diffèrent des modèles d'accumulation de type Ad Hoc par trois aspects :

- L'exposition est conçue comme un phénomène stochastique et non pas déterministe

- Les hypothèses explicites qui sont faites quant aux comportements de lecture ou d'écoute sont vérifiables par des tests statistiques.

- Il est possible d'estimer simultanément la couverture nette cumulée et la distribution des occasions de voir tandis que les modèles du type Ad Hoc sont limités à l'estimation de la couverture nette cumulée.

Nous étudierons tout d'abord les modèles d'ordre zéro en commençant par le modèle homogène suivi par les divers modèles hétérogènes. Ensuite, nous présenterons un modèle homogène d'ordre un, puis nous terminerons l'étude par l'examen des divers modèles de simulation.

Modèles d'accumulation d'ordre zéro.

321 . Le modèle d'accumulation binomial.

Hypothèses :

a) - La population est homogène. Chaque individu a la même probabilité d'exposition à une parution donnée d'un support donné.

b) - Les probabilités d'exposition sont stationnaires d'une parution à une autre. Ainsi, toutes les parutions d'un support donné sont considérés comme identiques.

Nous utiliserons les notations suivantes :

Soit I, l'ensemble des individus considérés : $I = \{1, 2, \ldots, i, \ldots, POP\}$
Soit N, l'ensemble des parutions considérées : $N = \{1, 2, \ldots, n, \ldots, N\}$

Soit P_{in}, la probabilité d'exposition du ième individu à la nième parution. En vertu des deux hypothèses précédentes, nous avons :

$$P_{in} = P_{i'n'} = p \qquad \begin{cases} \forall\, i, i' \in I \;;\; i \neq i' \\ \forall\, n, n' \in N \;;\; n \neq n' \end{cases}$$

La valeur de p, qui représente la probabilité d'exposition au support quel que soit l'individu et quelle que soit la parution, peut être estimée en

divisant l'audience moyenne du numéro du support par la population totale considérée. En vertu des deux hypothèses du modèle, l'exposition au numéro est une variable de Bernou lli. En conséquence, la probabilité qu'un individu soit exposé à exactement K parutions parmi N est donnée par la loi binomiale :

$$(21) \quad P(K / N, p) = \binom{N}{K} p^K (1 - p)^{N - K} \qquad 0 \leq K \leq N$$

L'espérance du nombre d'individus exposés à exactement K parutions parmi N est donnée par :

$$(22) \quad E_K^N = POP * P(K / N, p)$$

La probabilité qu'un individu soit exposé à, au moins, K parutions parmi N est donnée par :

$$(23) \quad P(x \geq K / N, P) = \Sigma_{x = k}^{N} \binom{N}{x} p^x (1 - p)^{N - x}$$

En particulier, la probabilité qu'un individu soit exposé à, au moins, une parution pârmi N, c'est-à-dire la probabilité qu'un individu fasse partie de la couverture nette générée par N insertions dans le support est de :

$$P(x \geq 1 / N, p) = 1 - P(x = 0/N, p) = 1 - (1 - p)^N$$

Puisque tous les individus sont par hypothèse identiques, l'espérance du nombre d'individus couverts par N parutions est de :

$$(24) \quad R_N = POP * \left[1 - (1 - p)^N \right]$$

Finalement, la probabilité que N parutions soient nécessaires pour atteindre un individu exactement K fois est donnée par la distribution de Pascal :

$$(25) \quad P(N / K, p) = \binom{N - 1}{K - 1} p^K (1 - p)^{N - K}$$

Le modèle binomial est le modèle stochastique d'accumulation le plus simple qui puisse être construit. De toute évidence, cette simplicité n'est acquise qu'au prix du recours à deux hypothèses fort peu réalistes. Par conséquent, nous ne pouvons pas nous attendre à ce que ce modèle s'ajuste de façon satisfaisante aux données disponibles. La principale faiblesse du modèle tient à ce que

les lecteurs d'une parution donnée ne sont pas des lecteurs plus probables de la prochaine parution que les non lecteurs de la première parution. De fait, les données dont nous disposons montrent que les audiences ne s'accumulent pas au hasard, comme prévu par le modèle binomial, mais à un taux beaucoup plus lent. Les deux tables suivantes illustrent clairement ce phénomène en comparant les prévisions du modèle binomial avec la couverture nette réelle observée pour 10 magazines.

Le modèle binomial surestime systématiquement et très gravement la couverture nette cumulée. L'erreur type moyenne de prédiction est de 0,185. Le modèle binomial est clairement inadéquat, spécialement eut regard à la qualité des prévisions fournies par certains modèles d'accumulation de type Ad Hoc, tel que le modèle logarithmique ou le modèle d'Agostini. Cependant, le modèle binomial présente quelque utilité en tant que point de repère auquel comparer les résultats des modèles stochastiques plus sophistiqués. Il sera aussi utile lorsque l'on cherchera à estimer la couverture nette cumulée de supports dont on ne connaît rien d'autre que l'audience moyenne d'une parution.

Finalement, le modèle binomial permet d'estimer directement la distribution des occasions de voir, ce que ne permettent pas les modèles de type Ad-Hoc. Toutefois, ces estimations sont bien loin d'être satisfaisantes comme le montre la table suivante qui les compare à la distribution des occasions de voir obtenue par conversion hypergéométrique inverse de la couverture nette cumulée. On remarquera que la distribution des fréquences binomiales est en forme de cloche, le mode se situant au niveau d'une exposition, tandis que les fréquences réelles suivent une courbe en U. Pour les dix magazines, la plus forte fréquence se situe au niveau de la non exposition, puis elle décroît régulièrement pour atteindre un minimum au niveau de 4 ou 5 expositions pour croître ensuite légèrement au niveau de 6 expositions, à cause des lecteurs fidèles. Cette forme très typique de la distribution des occasions de voir tient au fait bien connu que les individus sont, soit des lecteurs réguliers, soit des non lecteurs du magazine considéré et qu'il y a relativement peu de lecteurs occasionnels. A ce point de vue, le modèle binomial échoue complètement puisqu'il prédit une majorité de lecteurs occasionnels.

Le modèle binomial d'accumulation

Estimation de la couverture nette

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
1	22,1	22,1	11,8	11,8	15,1	15,1	9,6	9,6	19,2	19,2
2	32,4	39,3	17,8	22,2	23,9	27,9	14,4	18,3	25,9	34,7
3	39,1	52,7	21,8	31,4	30,0	38,8	17,7	26,1	29,6	47,2
4	44,0	63,2	25,0	39,5	34,8	48,0	20,2	33,2	32,0	57,4
5	47,7	71,3	27,5	46,6	38,6	55,9	22,2	39,6	33,8	65,6
6	50,6	77,7	29,6	52,9	41,8	62,6	23,9	45,4	35,3	72,2
Ecart type	0,185		0,143		0,129		0,130		0,238	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
1	31,2	31,2	19,4	19,4	23,8	23,8	19,4	19,4	36,0	36,0
2	42,6	52,7	27,4	35,0	34,4	41,9	27,3	35,0	46,5	59,0
3	49,4	67,4	32,6	47,6	41,2	55,7	32,2	47,6	52,5	73,8
4	54,1	77,6	36,5	57,8	46,0	66,3	35,6	57,8	56,6	83,2
5	57,7	84,6	39,6	66,0	49,7	74,3	38,3	66,0	59,7	89,3
6	60,5	89,4	42,1	72,6	52,7	80,4	40,4	72,6	62,2	93,1
Ecart type	0,206		0,198		0,185		0,208		0,229	

Le modèle binomial d'accumulation

Estimation de la distribution des occasions de voir

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Exposit.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
0	49,4	22,3	70,4	47,1	58,2	37,4	76,1	54,6	64,7	27,8
1	17,4	38,0	12,6	37,8	19,2	40,0	10,2	34,8	9,0	39,7
2	12,0	27,0	6,8	12,6	9,8	17,8	4,5	9,2	4,5	23,6
3	8,0	10,2	3,0	2,3	5,0	4,2	4,0	1,3	6,0	7,5
4	3,0	2,2	1,5	0,2	3,0	0,6	1,5	0,1	6,0	1,3
5	4,2	0,3	1,5	0,0	2,2	0,0	1,8	0,0	3,6	0,1
6	6,0	0,0	4,2	0,0	2,6	0,0	1,9	0,0	6,2	0,0
Chi deux	68,4		70,0		43,3		77,1		220,5	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Exposit.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
0	39,5	10,6	57,9	27,4	47,3	19,6	59,6	27,3	37,8	6,9
1	16,7	28,9	15,1	39,6	17,8	36,7	12,4	39,6	15,0	23,2
2	11,0	32,7	8,2	23,8	11,3	28,6	8,6	23,9	9,5	32,6
3	9,4	19,8	4,9	7,7	7,3	11,9	4,9	7,7	6,9	24,4
4	6,7	6,7	5,2	1,4	6,4	2,8	5,2	1,4	8,5	10,3
5	6,6	1,2	2,5	0,1	4,4	0,4	4,0	0,1	6,8	2,3
6	10,1	0,1	6,2	0,1	5,5	0,0	5,3	0,0	15,5	0,2
Chi deux	98,7		98,1		76,8		117,9		148,6	

Les valeurs élevées du test du chi deux démontrent la pauvreté de l'ajustement obtenu par le modèle binomial. Ce point méritait d'être souligné, étant donné que de nombreux chercheurs ont succombé aux simplifications tentantes offertes par le modèle binomial d'accumulation d'audience (14).

La prochaine catégorie de modèle que nous allons étudier rejette l'hypothèse d'homogénéité de la population. Ceci peut s'effectuer de trois manières :

- Premièrement, on peut admettre que la population cible est composée de divers segments, chacun d'eux étant caractérisé par une probabilité d'exposition spécifique. Cette approche a l'avantage de la simplicité, mais elle suppose implicitement que les individus composant un segment sont parfaitement homogènes quant à leurs probabilités d'exposition.

- Deuxièmement, on peut admettre que chaque individu faisant partie de la population cible est caractérisé par sa propre probabilité personnelle d'exposition. Cette approche, plus générale que la précédente, permet une représentation très flexible de l'hétérogénéité de la population. Il suffira d'admettre que les probabilités individuelles d'exposition suivent une certaine fonction de répartition caractérisant le degré d'hétérogénéité de la population cible.

- Troisièmement, on peut construire des modèles mixtes caractérisés par une double hétérogénéité. On admet alors que la population cible est composée de divers segments, chacun d'eux étant caractérisé par sa propre distribution des probabilités individuelles d'exposition.

La première option, qui n'introduit l'hétérogénéité qu'au niveau des segments, a été utilisée par Lee et Burkhart (15), Lee (16), Taylor (17), Ellis (18), Agostini (19) et Frankel (20). La seconde option, qui introduit l'hétérogénéité du niveau des individus, a été proposée par Hyett (21), Politz (22), Simmons (23), Greene et Stock (2), Greene (24), Friedman (25) et Chandon (26). La troisième option, qui introduit une double hétérogénéité, a été proposée par Broadbent (27), Morgensztern (28) et Chandon (26).

322. Les Modèles d'Accumulation à population segmentée

Hypothèses :

a- La population est hétérogène et comprend plusieurs segments. Chaque segment est homogène et tous les individus qui le composent ont la même probabilité d'exposition au support.

b- Les probabilités d'exposition caractérisant chaque segment sont stationnaires d'une parution à une autre. Les probabilités d'exposition de chaque segment sont indépendantes les unes des autres.

Le modèle d'accumulation le plus simple de cette catégorie ne distingue que deux segments :

- les lecteurs (ou auditeurs) qui représentent un proportion "a" de la population cible et qui sont caractérisés par une probabilité d'exposition "p" non nulle.

- les non lecteurs (ou auditeurs) qui représentent une proportion 1-a de la population et dont la probabilité d'exposition est nulle

Sous ces hypothèses, la probabilité qu'un lecteur (ou auditeur) soit exposé à exactement K parutions est donnée par la loi binomiale. Puisque tous les lecteurs ont la même probabilité "p" d'exposition, l'espérance du nombre de lecteurs exposés à K parutions parmi N est donnée par :

$$(26) \quad \begin{cases} E_K^N = a * POP * \binom{N}{K} p^K (1-p)^{N-K} & \text{avec } 1 \leq K \leq N \\ E_0^N = \left(1 - a + a (1-p)^N\right) & \text{avec } K = 0 \end{cases}$$

La probabilité qu'un lecteur soit exposé à, au moins, une parution parmi N est $1 - (1-p)^N$ donc, l'espérance de couverture nette après N parutions est de ;

$$(27) \qquad R_N = a * POP \times \left(1 - (1-p)^N\right)$$

Puisque le modèle d'accumulation à deux segments ne dépend que des deux paramètres a et p, ceux-ci peuvent être estimés à partir de l'audience des deux premières parutions d'un support.

Par application de l'équation (26), nous avons :

$$\begin{cases} R_1 = a * POP * p \\ R_2 = a * POP * \left(1 - (1-p)^2\right) \end{cases}$$

En résolvant ces deux équations nous obtenons les valeurs estimées des paramètres a et p.

$$(28) \begin{cases} \hat{p} = (2\ R_1 - R_2)\ /\ R_1 \\ \hat{a} = R_1^2\ /\ \left(POP * (2\ R_1 - R_2)\right) \end{cases}$$

Ce modèle assez simple procure une amélioration sensible par rapport au modèle binomial comme le montrent les deux tables suivantes qui comparent les estimations du modèle d'accumulation à deux segments à la couverture nette réelle et à la distribution des occasions de voir pour 10 magazines.

Toutefois, les estimations obtenues pour le paramètre "a" sont clairement inacceptables pour les dix magazines. Nous observons en effet que la proportion estimée de lecteurs est dans tous les cas inférieure à la couverture nette observée pour six parutions. Cette impossibilité logique est causée par l'hypothèse d'homogénéité des probabilités d'exposition à l'intérieur du segment des lecteurs. Cette hypothèse n'est pas réaliste, spécialement lorsque l'on considère que les lecteurs ne forment qu'un seul segment.

Le Modèle d'Accumulation à deux segments

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	39,1	37,2	21,5	20,9	30,0	29,0	17,7	16,8	29,6	28,2
4	44,0	39,4	25,0	22,4	34,8	32,0	20,2	18,0	32,0	29,1
5	47,7	40,5	27,5	23,2	38,6	33,8	22,2	18,6	33,8	29,3
6	50,6	41,0	29,6	23,6	41,8	34,8	23,9	18,9	35,3	29,4
$\hat{a}$	0,414		0,240		0,362		0,192		0,295	
$\hat{p}$	0,534		0,492		0,417		0,500		0,651	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	49,4	46,8	32,6	30,7	41,2	39,1	32,2	30,5	52,5	49,6
4	54,1	48,3	36,5	32,1	46,0	41,2	35,6	31,8	56,6	50,5
5	57,7	48,8	39,6	32,6	49,7	42,2	38,3	32,4	59,7	50,7
6	60,5	49,0	42,1	39,2	52,7	42,6	40,4	32,6	62,2	50,8
$\hat{a}$	0,492		0,330		0,429		0,327		0,508	
$\hat{p}$	0,635		0,588		0,555		0,593		0,708	

Le Modèle d'Accumulation à Deux Segments

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	59,0	70,4	76,4	58,2	65,2	76,1	81,1	64,7	70,6
1	17,4	2,9	12,6	2,4	19,2	6,1	10,2	1,8	9,0	0,6
2	12,0	8,4	6,8	5,8	9,8	10,9	4,5	4,5	4,5	2,8
3	8,0	12,8	3,0	7,5	5,0	10,4	4,0	6,0	6,0	6,9
4	3,0	10,9	1,5	5,4	3,0	5,6	1,5	4,5	6,0	9,7
5	4,2	5,0	1,5	2,1	2,2	1,4	1,8	1,8	3,6	7,2
6	6,0	1,0	4,2	0,4	2,6	0,4	1,9	0,3	6,2	2,2
Chi Deux	43,1		29,5		20,1		15,6		17,62	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	51,0	57,9	67,1	47,3	57,4	59,6	67,4	37,8	49,2
1	16,7	1,4	15,1	1,4	17,8	2,5	12,4	1,3	15,0	0,5
2	11,0	5,3	8,2	4,9	11,3	7,8	8,6	4,7	9,5	2,8
3	9,4	12,3	4,9	9,4	7,3	12,9	4,9	9,2	6,9	9,0
4	6,7	15,6	5,2	10,1	6,4	12,1	5,2	10,1	8,5	16,3
5	6,6	11,1	2,5	5,7	4,4	6,0	4,0	5,9	6,8	15,8
6	10,1	3,3	6,2	1,4	5,5	1,3	5,3	1,4	15,5	6,4
Chi Deux	40,7		31,8		29,6		24,9		47,2	

Toutefois, le modèle d'accumulation à population segmentée n'est pas nécessairement limité à la considération de deux segments ainsi que dans le modèle d'Ellis (18). Agostini préconise l'emploi de trois ou quatre segments(19), tandis que Frankel (20) propose l'emploi de cinq segments. A mesure que le nombre de segments considérés augmente, l'hétérogénéité de la population est mieux prise en considération et l'ajustement aux données s'améliore. Cependant, en contrepartie, le nombre de paramètres s'accroît et il faut connaître la couverture nette cumulée d'un plus grand nombre de parutions pour pouvoir les estimer. Par exemple, avec trois segments, nous devons connaître la couverture nette cumulée de quatre parutions, ou bien alternativement, la distribution des occasions de voir sur quatre parutions. Puisque ces données sont coûteuses et difficiles à obtenir par la méthode de la lecture au numéro, il n'est pas surprenant de trouver parmi les avocats du modèle d'accumulation multisegments les défenseurs de la méthode des habitudes de lecture. Afin d'illustrer le fonctionnement du modèle multisegments, supposons que la population se compose de trois segments: les lecteurs réguliers, les lecteurs occasionnels et les non-lecteurs. Si nous ne posons pas d'hypothèses supplémentaires, nous devons estimer les quatre paramètres suivants :

a = la proportion de lecteurs réguliers

b = la proportion de lecteurs occasionnels

p_1 = la probabilité d'exposition parmi les lecteurs réguliers

p_2 = la probabilité d'exposition parmi les lecteurs occasionnels

La proportion de non-lecteurs sera donnée par $c = 1 - a - b$ et, leur probabilité d'exposition est, par définition, nulle.

Les paramètres a, b, p_1, et p_2 peuvent être estimés soit à partir des équations de la couverture nette, soit à partir des équations de la distribution des occasions de voir.

$$\begin{cases} R_1 = a\,p_1 + b\,p_2 \\ R_2 = a\,\left(1-(1-p_1)^2\right) + b\,\left(1-(1-p_2)^2\right) \\ R_3 = a\,\left(1-(1-p_1)^3\right) + b\,\left(1-(1-p_2)^3\right) \\ R_4 = a\,\left(1-(1-p_1)^4\right) + b\,\left(1-(1-p_2)^4\right) \end{cases}$$

$$\begin{cases} E_4^4 = a\,p_1^4 + 4\,p_2^4 \\ E_3^4 = 4\,a\,p_1^3\,(1-p_1) + 4\,b\,p_2^3\,(1-p_2) \\ E_2^4 = 6\,a\,p_1^2\,(1-p_1)^2 + 6\,b\,p_2^2\,(1-p_2)^2 \\ E_1^4 = 4\,a\,p_1\,(1-p_1)^3 + 4\,b\,p_2\,(1-p_2)^3 \end{cases}$$

Quelle que soit la solution choisie, l'estimation des quatre paramètres est malaisée puisqu'elle requiert la résolution d'un système non linéaire de quatre équations. De plus, l'unicité de la solution n'est pas garantie. Pour pallier à cette difficulté, diverses méthodes ont été proposées. Une solution consiste à admettre que tous les individus ayant déclaré être exposés en moyenne à K parutions sur N possèdent une probabilité d'exposition de K/N. Cette méthode, dite de l'affectation actuarielle des probabilités, conduit à des résultats erronés comme l'ont démontré Simmons (29), Greene (24) et Schreiber (30). D'autres méthodes permettent de résoudre ce problème. Nous les examinerons en détail, en fin de chapitre, lorsque nous traiterons des modèles de simulation d'audience. Nous allons présenter ici une méthode simple et nouvelle qui permet d'estimer les paramètres d'un modèle à 3 segments sur la seule base de la couverture nette cumulée des deux premières parutions.

Supposons que l'audience se compose de 3 segments, définis comme suit :

Segments	Probabilité	Proportion
Lecteurs réguliers	p_1	a
Lecteurs occasionnels	p_2	b
Non lecteurs	0	1 - a - b

Ecrivons les équations de la couverture nette pour une et deux parutions :

$$\begin{cases} R_1 = a\,p_1 + b\,p_2 \\ R_2 = a\,p_1\,(2-p_1) + b\,p_2\,(2-p_2) \end{cases}$$

En résolvant ces deux équations pour a et b, nous obtenons :

$$(29)\begin{cases} \hat{a} = \left((2-p_2)\,R_1 - R_2\right) \,/\, \left(p_1\,(p_1-p_2)\right) \\ \hat{b} = \left(R_2 - (2-p_1)\,R_1\right) \,/\, \left(p_2\,(p_1-p_2)\right) \end{cases}$$

Puisque a et b doivent être positifs, le modèle n'est applicable que si les conditions suivantes sont vérifiées par p_1 et p_2 :

$$(30) \quad (2-p_1)\,R_1 < R_2 < (2-p_2)\,R_1$$

Ces conditions nous permettent de définir des couples de probabilités p_1 et p_2 qui vérifient la double inégalité (30) :

$$(31) \qquad p_2 < 2 - (R_2/R_1) < p_1$$

Parmi tous les couples de probabilités vérifiant (31), on choisira le couple centré sur $2 - (R_2/R_1)$ tel que l'intervalle (p_2, p_1) corresponde à une valeur choisie arbitrairement pour distinguer commodément les lecteurs réguliers des lecteurs occasionnels.

Le Modèle d'Accumulation à Trois Segments

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	39,1	39,2	21,8	21,9	30,0	30,4	17,7	17,7	29,6	30,0
4	44,0	44,2	25,0	25,1	34,8	35,7	20,2	20,2	32,0	32,6
5	47,7	48,0	27,5	27,6	38,6	38,6	22,2	22,2	33,8	34,3
6	50,6	50,9	29,6	29,7	41,8	44,4	23,9	23,7	35,3	35,4
% Réguliers	13,3		7,5		10,5		6,0		10,1	
% Occasion.	47,2		30,8		64,4		24,0		27,4	
Erreur type	0,0026		0,0010		0,0168		0,0010		0,0043	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	49,4	49,6	32,6	32,4	41,2	41,3	32,2	32,3	52,5	52,1
4	54,1	54,2	36,5	36,0	46,0	46,2	35,6	35,7	56,6	55,2
5	57,7	57,2	39,6	38,5	49,7	49,1	38,3	38,1	59,7	57,0
6	60,5	59,3	42,1	40,2	52,7	52,6	40,4	39,9	62,2	58,0
% Réguliers	16,7		10,9		13,9		10,9		18,1	
% Occasion.	46,6		33,7		46,7		33,1		46,1	
Erreur Type	0,0067		0,0112		0,0016		0,0029		0,0251	

Le Modèle d'Accumulation à Trois Segments

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	49,1	70,4	70,3	58,2	55,6	76,1	76,3	64,7	64,6
1	17,4	17,5	12,6	12,2	19,2	24,3	10,2	9,4	9,0	6,6
2	12,0	13,4	6,8	7,4	9,8	8,6	4,5	6,0	4,5	9,0
3	8,0	6,1	3,0	3,0	5,0	3,2	4,0	2,5	6,0	6,5
4	3,0	3,9	1,5	2,3	3,0	3,5	1,5	1,8	6,0	2,9
5	4,2	5,5	1,5	2,2	2,2	3,4	1,8	2,4	3,6	2,9
6	6,0	4,5	4,2	1,8	2,6	1,4	1,9	1,6	6,2	7,5
Chi Deux	1,7		3,2		3,6		1,5		7,1	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	40,7	57,9	59,8	47,3	47,4	59,6	60,1	37,8	42,0
1	16,7	12,2	15,1	10,7	17,8	16,4	12,4	10,3	15,0	5,9
2	11,0	15,4	8,2	10,8	11,3	14,1	8,6	10,7	9,5	11,6
3	9,4	10,4	4,9	6,0	7,3	6,9	4,9	6,1	6,9	12,2
4	6,7	4,7	5,2	3,0	6,4	4,0	5,2	3,0	8,5	7,4
5	6,6	5,4	2,5	4,3	4,4	5,8	4,0	4,3	6,8	4,8
6	10,1	11,2	6,2	5,4	5,5	5,4	5,3	5,5	15,5	16,1
Chi Deux	4,0		4,8		2,1		2,1		11,4	

Les deux tables précédentes, décrivent les résultats obtenus en utilisant l'intervalle : $P_2 - P_1 = 0,6$ pour chacun des 10 magazines.

L'erreur type moyenne de prédiction est 0,0073 pour la couverture nette cumulée. Le modèle d'accumulation à 3 segments est donc nettement supérieur au modèle d'accumulation à 2 segments, tout en étant légèrement inférieur au modèle d'Agostini.

En ce qui concerne la distribution des occasions de voir, l'ajustement du modèle est satisfaisant dans 4 cas sur 10. On observe en effet quatre chi deux inférieurs à la valeur 2,2 qui apparaîtrait 10 fois sur 100 si les distributions observées et estimées étaient en fait tirées de la même distribution.

En conclusion, les modèles d'accumulation qui introduisent l'hétérogénéité de la population au niveau des segments ne donnent des résultats satisfaisants que si trois segments au moins sont considérés. Si l'on souhaite introduire un plus grand nombre de segments il faut alors connaître la distribution des occasions de voir pour quatre parutions, au moins. Comme ces données ne sont pas disponibles par la technique de la double enquête, nous avons développé et testé un modèle simple, à trois segments, qui ne requiert que la couverture nette cumulée des deux premières parutions.

Toutefois, il n'existe aucune base théorique ou expérimentale permettant de soutenir l'hypothèse de l'homogénéité des individus à l'intérieur des segments. Nous allons donc étudier maintenant les divers modèles qui permettent de prendre en compte les différences individuelles dans les probabilités d'exposition.

323. Les Modèles d'accumulation à hétérogénéité individuelle

Les divers modèles groupés sous ce titre partagent ensemble les deux hypothèses suivantes .

Hypothèses :

a- La population cible est hétérogène. Chaque individu est caractérisé par sa propre probabilité personnelle d'exposition à un support donné. Ces probabilités personnelles d'exposition possèdent une distribution continue dans la population cible.

b- Les probabilités personnelles d'exposition sont stationnaires d'une parution à une autre, ce qui implique que pour tout individu toutes les parutions sont équiprobables. Les probabilités d'exposition sont stationnaires d'un individu à un autre, ce qui implique que la probabilité d'exposition d'un individu est indépendante des probabilités d'exposition des autres individus.

Les modèles d'accumulation à hétérogénéité individuelle ne diffèrent entre eux que sur un point crucial, le choix de la distribution des individus le long du continuum des probabilités d'exposition. Si cette distribution est elle-même une densité de probabilité, alors nous avons affaire à ce que les statisticiens appellent une distribution composée. Les analyses qui vont suivre sont facilitées par l'adoption de la notation suivante (31) :

pi = probabilité personnelle d'exposition de l'individu à une parution quelconque du support considéré. L'indice i sera omis lorsqu'aucun risque de confusion n'est possible.

POP = Nombre total d'individus inclus dans la population cible.

f(p)= Nombre d'individus possédant la probabilité d'exposition p.

E_K^N = Nombre d'individus exposés exactement à K parutions parmi N

R_N = Nombre d'individus exposés, au moins, à une parution parmi N.

Par souci de simplicité, la distribution E_K^N et la couverture nette cumulée seront exprimées en pourcentage de la population ainsi qu'en valeur absolue sans changement de notation.

Le degré et la nature de l'hétérogénéité de la population est défini par la spécification de f(p). En fait, nous avons déjà examiné le cas où f(p) est une fonction discrète, lorsque nous avons traité des modèles d'accumulation à population segmentée. Nous nous restreindrons donc désormais au cas où f(p) est une fonction continue. Deux possibilités doivent être distinguées. Tout d'abord, f(p) peut être une fonction continue quelconque, choisie sur la base de considérations empiriques comme c'est le cas pour le modèle d'accumulation de Politz. Dans ce cas, la distribution des occasions de voir E_K^N est une distribution composite. Si, par contre, f(p) est une densité de probabilité, alors la distribution des occasions de voir E_K^N est une distribution composée.

Il est particulièrement important de ne pas confondre f(p), qui mesure l'hétérogénéité de la population du point de vue des probabilités d'exposition, avec E_K^N qui mesure la distribution des occasions de voir. La fonction f(p) est inconnue, elle n'est pas directement observable et elle est continue. La distribution E_K^N est par contre observable et discrète. La modélisation des E_K^N comprend quatre étapes. Premièrement, la formulation d'une hypothèse quant à l'équation de f(p), deuxièmement la déduction à partir de cette hypothèse de l'équation des E_K^N, troisièmement l'estimation des paramètres de la distribution des E_K^N et quatrièmement le test de la qualité de l'ajustement du modèle aux données.

Analysons tout d'abord la relation mathématique qui lie f(p) à la distribution des occasions de voir E_K^N et à la couverture nette cumulée R_N.

En premier lieu, l'audience cumulée de N parutions est égale, d'une part, à la population totale moins les individus non exposés, et, d'autre part, à la somme, pour toutes les classes de fréquence, des individus exposés.

$$R_N = POP - E_0^N = \Sigma_{K=1}^N E_K^N$$

En second lieu, la population totale est égale, d'une part, à la somme des individus associés à chacune des valeurs de p et, d'autre part, à la somme de la distribution des occasions de voir.

$$POP = \int_0^1 f(p)dp = \Sigma_{K=0}^N E_K^N$$

Soit un individu i et sa probabilité personnelle d'exposition p_i. En vertu de l'hypothèse b, l'exposition à une parution quelconque du support est

une variable de Bernoulli et, par conséquent, la probabilité que cet individu i soit exposé à exactement K parutions parmi N est donnée par la distribution binomiale :

$$P(K/N,Pi) = \binom{N}{K} Pi^{K}(1-Pi)^{N-K} \qquad 0 \leq K \leq N$$

Le nombre d'individus qui, comme l'individu i, possèdent la probabilité d'exposition p_i est (f(pi). Donc, l'espérance du nombre d'individus exposés à exactement K parutions parmi N, parmi les individus possédant la probabilité pi, est donné par :

$$i\ E_K^N = \binom{N}{K} pi^K (1-pi)^{N-K} f(pi)$$

En sommant les individus pour toutes les probabilités pi possibles, nous obtenons :

$$(32) \qquad E_K^N = \int_0^1 \binom{N}{K} p^K(1-p)^{N-K} f(p)\, dp$$

Cette équation qui lie E_K^N à f(p), a été présentée pour la première fois par Politz en 1958. A partir de cette équation et en utilisant la formule de conversion (1) présentée au début du chapitre, il est possible de calculer la couverture nette cumulée. De même, en utilisant la formule de conversion (4), on peut calculer la distribution des audiences fidèles.

Nous allons démontrer l'existence d'une relation de récurrence qui permet de calculer la distribution des occasions de voir pour N-1 parutions à partir de la distribution des occasions de voir pour N parutions. En outre, cette relation reste vraie quelle que soit la forme de f(p). Réécrivons l'équation (32) en mettant en facteur (1-p) et le coefficient binomial.

$$E_K^N = \binom{N}{K} \int_0^1 p^K(1-p)^{N-K-1} (1-p)\, f(p)\, dp$$

En effectuant le produit par (1-p) sous le signe intégral, nous obtenons :

$$p^K(1-p)^{N-K-1}(1-p) = p^K(1-p)^{N-K-1} - p^{K+1}(1-p)^{N-K-1}$$

En divisant des deux côtés du signe égal par $\binom{N}{K}$, il vient :

$$E_K^N/\binom{N}{K} = \int_0^1 p^K(1-p)^{N-K-1} f(p)dp - \int_0^1 p^{K+1}(1-p)^{N-K-1} f(p)\, dp$$

En vertu de l'équation (32), les deux expressions à droite du signe égal sont respectivement égales à : $E_K^{N-1} / \binom{N-1}{K}$ et à : $E_{K+1}^N / \binom{N}{K+1}$

En réarrangeant les termes et en simplifiant les coefficients binomiaux, nous obtenons la relation de récurrence suivante :

$$(33) \qquad E_K^{N-1} = \left(\frac{N-K}{N} \right) E_K^N + \left(\frac{K+1}{N} \right) E_{K+1}^N$$

Cette relation est comparable à la relation de récurrence du triangle de Pascal pour les coefficients binomiaux. Elle permet de calculer la distribution des occasions de voir pour tout nombre de parutions inférieur à N en fonction de la distribution des occasions de voir pour N parutions.

Pour la couverture nette, une relation équivalente à la relation (33) est facilement dérivée :

$$(34) \qquad R_{N-1} = R_N - E_1^N / N$$

Les deux équations de récurrence (33) et (35), qui n'apparaissent dans aucune recherche antérieure sur le sujet, sont extrêmement utiles. Elles permettent de reconstituer, dans sa totalité, l'ensemble des processus d'exposition qui aboutissent à la distribution des occasions de voir observée pour N parutions. La table suivante illustre cette reconstitution pour les magazines Life, Post, Ladies Home Journal, Look and Readers'Digest.

L'équation de récurrence (32) peut être généralisée au calcul direct de la distribution des occasions de voir pour N-t parutions en fonction de la distribution des occasions de voir pour N parutions.

$$E_K^{N-t} = \Sigma_{i=0}^{t} \left[\binom{t}{i} \binom{N-t}{K} / \binom{N}{K+i} \right] E_{K+i}^N$$

L'analyse des relations mathématiques liant l'hétérogénéité de la population à la distribution des occasions de voir et à la couverture nette cumulée étant terminée, nous allons examiner deux modèles d'accumulation qui ne diffèrent que par leur spécification de f(p).

Couverture Nette, Fréquence Moyenne et Distribution des Occasions de Voir

Parution	Couverture Nette	Fréquence Moyenne	Distribution des Occasions de voir 0	1	2	3	4	5	6
			Magazine Life, 1958						
1	15320	1,00	33820	15320					
2	20930	1,46	28210	11220	9710				
3	24290	1,89	24850	10080	6750	7460			
4	26610	2,30	22530	9280	6240	4840	6250		
5	28340	2,70	20800	8650	5900	4500	3800	5490	
6	29710	3,09	19430	8220	5400	4600	3300	3240	4950
			Magazine Post, 1958						
1	9530	1,00	39610	9530					
2	13450	1,42	35690	7840	5610				
3	16030	1,78	33110	7740	4020	4270			
4	17930	2,13	31210	7600	4080	2640	3610		
5	19440	2,45	29700	7550	3900	2900	1850	3240	
6	20680	2,76	28460	7440	4050	2400	2550	1200	3040
			Magazine Look, 1958						
1	11690	1,00	37450	11690					
2	16910	1,38	32230	10440	6470				
3	20230	1,73	28910	9960	5700	4570			
4	22610	2,07	26530	9520	5840	3840	3610		
5	24440	2,39	24700	9150	5500	3900	2850	3040	
6	25900	2,71	23240	8760	5550	3600	3150	2160	2680
			Magazine Ladies Home Journal, 1958						
1	9550	1,00	39590	9550					
2	13420	1,42	35720	7740	5680				
3	15810	1,81	35330	7170	4440	4200			
4	17510	2,18	31630	6800	4140	3160	3410		
5	18810	2,54	30330	6500	4000	2900	2500	2910	
6	19830	2,89	29310	6120	4200	2400	2550	1980	2580
			Magazine Readers'Digest, 1958						
1	17675	1,00	31465	17675					
2	22840	1,55	26300	10330	12510				
3	25790	2,06	23350	8850	6645	10295			
4	27810	2,54	21330	8080	5580	5140	9010		
5	29350	3,01	19790	7700	4800	4500	4175	8175	
6	30580	3,47	18560	7380	4650	3400	4200	3330	7620

Le Modèle d'Accumulation de Politz

Alfred Politz (22), dans son étude du magazine Life réalisée en 1953, propose d'approximer f(p) par un polynome du troisième degré :

$$f(p) = a + bp + cp^2 + dp^3$$

donc, Politz optait pour un modèle de type composite puisque f(p) n'est pas représenté par une densité de probabilité. En outre, puisque ce modèle dépend de quatre paramètres a,b,c,d, il ne peut être estimé que si la couverture nette cumulée ou la distribution des occasions de voir sont connues pour, au moins, quatre parutions. En fait, Politz estima les quatre paramètres sur la base de la distribution des occasions de voir pour six parutions. Ayant ainsi plus d'équations que d'inconnues, l'estimation des paramètres fut faite par la méthode des moindres carrés.

Plus tard, Charnes, Cooper, Learner et Snow (32) ont suggéré qu'une fonction du second degré pourrait suffire à l'estimation de f(p). Mais cela implique, néanmoins, de connaître la couverture nette cumulée ou la distribution des occasions de voir pour trois parutions.

Si l'on désire s'en tenir à la couverture nette de deux parutions, le seul polynome qui puisse être utilisé pour estimer f(p) est la ligne droite. Néanmoins, nous avons déjà remarqué dans nos analyses précédentes que l'audience des magazines tendait à se répartir aux deux extrêmes de l'échelle des probabilités d'exposition. Nous avons donc de bonnes raisons de croire qu'un modèle fondé sur une hétérogénéité linéaire donnera des résultats peu satisfaisants. Aussi, nous ne présentons ce modèle que dans le but de détailler le fonctionnement des modèles composites et de démontrer leur faiblesse sur le plan logique.

Supposons que f(p) soit linéaire : $f(p) = a + bp$

En vertu de l'équation (32), nous obtenons :

$$E_K^N = \int_0^1 \binom{N}{K} p^K(1-p)^{N-K}(a+bp)\, dp$$

$$E_K^N = \binom{N}{K} \left[a \int_0^1 p^K(1-p)^{N-K} dp + b \int_0^1 p^{K+1}(1-p)^{N-K}\, dp \right]$$

Les deux intégrales entre crochets sont des fonctions eulériennes de première espèce, encore appelées fonctions Bêta.

$$\beta(K+1\ ,\ N-K+1) = \int_0^1 p^K(1-p)^{N-K}\, dp$$

Nous pouvons donc réécrire E_K^N comme suit :

$$(35) \qquad E_K^N = \binom{N}{K} \left[a * \beta(K+1,N-K+1) + b * \beta(K+2,N-K+1) \right]$$

puisque l'enquête d'audience nous fournit les valeurs de E_1^2 et E_0^2, les paramètres a et b peuvent être estimés par les deux équations suivantes :

$$\begin{cases} E_1^2 = 2 \left[a * \beta(2,2) + b * \beta(3,2) \right] \\ E_0^2 = a * \beta(1,3) + b * \beta(2,3) \end{cases}$$

Etant donné que N et K sont des entiers, la fonction Bêta peut être calculée comme suit :

$$(36) \qquad \beta(K+1 , N-K+1) = K ! \ (N-K) ! \ / \ (N+1)!$$

Les paramètres a et b de la fonction linéaire d'hétérogénéité sont donc estimés par :

$$\begin{cases} \hat{a} = 6 E_0^2 - 3 E_1^2 \\ \hat{b} = 12 (E_1^2 - E_0^2) \end{cases} \quad \text{ou, en termes de couverture nette :} \quad \begin{cases} \hat{a} = 6 (1 + R_1 - 2R_2) \\ \hat{b} = 12 (3R_2 - 2R_1 - 1) \end{cases}$$

A partir des valeurs estimées de paramètres a et b, la distribution des occasions de voir pour un nombre quelconque de parutions s'obtient par l'équation (35). La couverture nette cumulée en pourcentage de la population totale s'obtient par :

$$R_N = 1 - \left(a * \beta(1,N+1) + b * \beta(2,N+1) \right)$$

Les deux tables suivantes comparent les estimations du modèle d'hétérogénéité linéaire à la couverture nette et à la distribution des occasions de voir observée pour 10 magazines.

Comme prévu, l'ajustement est très mauvais. L'erreur type de prédiction est en moyenne de 0,103. C'est le plus mauvais résultat parmi tous les modèles testés bien que l'amélioration soit sensible par rapport au modèle binomial. Mais il y a plus grave. Le paramètre b est supérieur en valeur absolue au paramètre a, pour tous les magazines. Cela implique que pour des niveaux de probabilité proche de 1, il existe un nombre négatif d'individus possédant une telle probabilité d'exposition ! Ceci est évidemment absurde et le modèle doit être rejeté pour ce motif.

Le Modèle d'Accumulation à Hétérogénéité Linéaire

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	39,1	42,2	21,8	27,8	30,00	34,1	17,7	24,4	29,6	35,3
4	44,0	50,0	25,0	36,6	34,8	42,6	20,2	33,4	32,0	43,4
5	47,7	56,1	27,5	43,9	38,6	49,4	22,2	40,9	33,8	49,9
6	50,6	60,9	29,6	49,7	41,8	54,8	23,9	47,0	35,3	55,2
$\hat{a}$	3,43		4,57		4,04		4,85		4,04	
$\hat{b}$	-5,64		-8,42		-7,02		-9,12		-7,28	
Erreur type	0,0750		0,1453		0,0956		0,1662		0,1432	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	49,4	51,8	32,6	37,1	41,2	44,1	32,2	36,9	52,5	55,0
4	54,1	58,6	36,5	45,1	46,0	51,8	35,6	45,0	56,6	61,4
5	57,7	63,9	39,6	51,6	49,7	57,7	38,3	51,5	59,7	66,3
6	60,5	67,9	42,1	56,7	52,7	62,4	40,4	56,6	62,2	70,1
$\hat{a}$	2,76		3,88		3,30		3,89		2,58	
$\hat{b}$	-4,15		-6,79		-5,33		-6,83		-3,90	
Erreur type	0,0549		0,1063		0,0706		0,1170		0,0581	

Le Modèle d'Accumulation a Hétérogénéité Linéaire

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	39,1	70,4	50,3	58,2	45,2	76,1	53,0	64,7	44,8
1	17,4	29,0	12,6	35,2	19,2	32,6	10,2	36,7	9,0	31,8
2	12,0	19,0	6,8	20,2	9,8	20,1	4,5	20,4	4,5	18,7
3	8,0	8,8	3,0	5,1	5,0	7,5	4,0	4,1	6,0	5,7
4	3,0	-1,2	1,5	-9,9	3,0	-5,0	1,5	-12,0	6,0	-7,3
5	4,2	-11,3	1,5	-24,9	2,2	-17,5	1,8	-28,5	3,6	-20,3
6	6,0	-21,4	4,2	-40,0	2,6	-30,1	1,9	-44,7	6,2	-33,3
Chi Deux	202,2		1091,9		632,9		1910,2		547,8	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	32,1	57,9	43,2	47,3	37,6	59,6	43,4	37,8	29,9
1	16,7	24,6	15,1	31,1	17,8	28,1	12,4	31,2	15,0	22,9
2	11,0	17,2	8,2	19,0	11,3	18,6	8,6	19,0	9,5	16,0
3	9,4	9,8	4,9	6,9	7,3	9,1	4,9	6,8	6,9	9,0
4	6,7	2,4	5,2	-5,3	6,4	-0,4	5,2	-5,4	8,5	2,0
5	6,6	-5,1	2,5	-17,4	4,4	-9,9	4,0	-17,6	6,8	-4,9
6	10,1	-12,1	6,2	-29,5	5,5	-19,5	5,3	-29,8	15,5	-11,9
Chi Deux	82,5		421,0		180,4		417,0		84,4	

Le Modèle d'Accumulation Bêta Binomial

G.P. Hyett (21) suggéra le premier, en 1958, l'utilisation de la distribution Bêta pour modeler l'hétérogénéité de la population. Plusieurs avantages découlent de l'emploi d'une véritable densité de probabilité, définie sur l'intervalle (0,1), pour représenter f(p).

a- Une densité de probabilité est positive pour tout son intervalle de définition, ainsi l'erreur logique d'un nombre négatif d'individus pour certains niveaux de probabilité est rendue impossible.

b- La surface comprise sous la courbe de densité de probabilité est égale à l'unité, donc l'erreur logique d'une couverture nette supérieure à l'unité ne peut se produire.

Ainsi, en représentant f(p) par une densité de probabilité définie sur l'intervalle (0,1) on prévient l'apparition de résultats illogiques ou absurdes. Cependant, les avantages obtenus en matière de robustesse ne sont pas les seuls à prendre en compte. On bénéficiera aussi du fait que la distribution des occasions de voir appartiendra à la famille des distributions binomiales composées. Ces distributions, qui ont été étudiées en statistique mathématique, possèdent un certain nombre de propriétés qui nous permettront de dériver plusieurs classes d'estimateur pour leurs paramètres. C'est ainsi que nous comparerons l'estimation par la méthode des moments, du maximum de vraisemblance, des moindres carrés et du chi deux.

Cependant, puisqu'il existe plusieurs densités de probabilité définies sur l'intervalle (0,1) pourquoi choisir la densité Bêta plutôt que toute autre ? Les raisons théoriques et pratiques qui justifient ce choix sont les suivantes :

1- La distribution Beta est la distribution conjuguée de la distribution binomiale. Ce qui signifie que, lorsque le paramètre d'un processus de Bernoulli est traité comme une variable aléatoire pi, la distribution a priori la plus commode des probabilités pi est la distribution Bêta, car la distribution postérieure résultant de la prise d'un échantillon sera encore une distribution Bêta.

2- La distribution Bêta est extrêmement flexible. Elle peut prendre presque toutes les formes possibles, concave vers le haut ou vers le bas, avec ou sans

point d'inflexion et même avec deux points d'inflexion. Elle peut être symétrique ou asymétrique, concentrée sur les faibles probabilités ou concentrée sur les fortes probabilités.

3- La distribution Bêta ne dépend que de deux paramètres A et B. Ainsi le modèle peut être ajusté à tout support, la couverture nette et la distribution des occasions de voir peuvent être extrapolées à tout nombre de parutions, sur la seule base de la couverture nette des deux premières parutions.

4- Lorsque l'on suppose que f(p) suit une densité Bêta, alors la distribution des occasions de voir suit une distribution composée simple et se prêtant bien aux calculs, la distribution Bêta Binomiale.

5- Finalement, nous allons montrer en détail que le choix de la densité Bêta conduit à une couverture nette cumulée et à une distribution des occasions de voir qui s'ajustent très bien aux données observables. Ce fait, antérieurement souligné par Greene et Stock (2) et par Simmons (29) a été si bien admis par les milieux publicitaires américains que, depuis 1964, les services de mesure d'audience s'en tiennent à l'estimation de la couverture nette cumulée de deux parutions, du moins dans le domaine des magazines.

Nous allons présenter maintenant une analyse mathématique détaillée du modèle d'accumulation Bêta Binomial, complétée par une discussion des diverses méthodes d'estimation de ses paramètres.

Supposons que f(p) suive une distribution Bêta de paramètres A et B :

$$(37) \quad f(p) = P(p/A,B) = \frac{\Gamma(A+B)}{\Gamma(A)\,\Gamma(B)}\, p^{A-1}(1-p)^{B-1}$$

avec $A > 0,\ B > 0\ , \qquad 0 \leq p \leq 1$

Le symbole Γ désigne la fonction gamma définie par :

$$\Gamma(A) = (A-1)\Gamma(A-1) = \int_0^\infty e^{-x}\, x^{A-1} dx$$

En remplaçant dans l'équation (32) f(p) par sa valeur donnée en (37), nous obtenons :

$$E_K^N = \binom{N}{K} \frac{\Gamma(A+B)}{\Gamma(A)\ \Gamma(B)} \int_0^1 p^K (1-p)^{N-K}\ p^{A-1}(1-p)^{B-1} dp$$

$$E_K^N = \binom{N}{K} \frac{\Gamma(A+B)}{\Gamma(A)\ \Gamma(B)} \int_0^1 p^{A+K-1}(1-p)^{N+B-K-1} dp$$

(38) $$E_K^N = \binom{N}{K} \frac{\Gamma(A+B)\Gamma(A+K)\Gamma(N+B-K)}{\Gamma(A)\Gamma(B)\ \Gamma(A+B+N)}$$ $0 \leq K \leq N$; entiers

$A,B \geq 0$; réels

L'expression (38) donne le pourcentage d'individus exposés à exactement K parutions parmi N, lorsque les probabilités individuelles d'exposition suivent une distribution Bêta de paramètres A et B.

En d'autres termes, l'équation (38) donne la probabilité d'obtenir exactement K succès en N essais lorsque les probabilités de réussite sont distribuées selon une distribution Bêta de paramètres A et B. Cette distribution est connue sous le nom de distribution Bêta Binomiale (DBB) (33).

Une expression rationnelle de E_K^N peut être dérivée de l'expression (38) afin de faciliter les calculs :

(39) $$E_K^N = \binom{N}{K} \frac{(A+K-1)(A+K-2)\ldots(A)(B+N-K-1)(B+N-K-2)\ldots.(B)}{(A+B+N-1)(A+B+N-2)\ldots.(A+B)}$$

Cette expression se simplifie davantage lorsque l'on désire calculer le pourcentage de non lecteur (i.e.. K=0)

$$E_0^N = \frac{(B+N-1)(B+N-2)\ldots.(B)}{(A+B+N-1)\ \ldots..\ (A+B)}$$

Ainsi, il apparaît que pour une valeur donnée de A, l'accroissement de B entraîne l'accroissement du pourcentage de non lecteurs. De même, lorsque K=N, l'expression (39) se réduit à :

$$E_N^N = \frac{(A+N-1)(A+N-2)\ \ldots.\ (A)}{(A+B+N-1)\ \ldots\ldots\ (A+B)}$$

Nous voyons donc que pour une valeur donnée de B, l'accroissement de A entraîne l'accroissement du pourcentage de lecteurs fidèles.

Un cas très particulier se produit lorsque A=B=1. La population est alors répartie de façon égale parmi tous les niveaux de probabilité d'exposition possible et la distribution des occasions de voir se réduit à $E_K^N = 1/(N+1)$ pour toutes les valeurs de K et de N.

L'audience moyenne d'une parution, exprimée en pourcentage de la population est égale au rapport du paramètre A à la somme des paramètres A+B. En effet nous avons :

$$E_1^1 = R_1 = A / (A+B)$$

Ainsi plus A est grand par rapport à B, plus l'audience moyenne d'une parution est élevée.

A partir de l'expression (39), la formule de la couverture nette est aisément établie :

$$(40) \qquad R_N = 1 - \frac{(B+N-1)(B+N-2)\ldots.(B)}{(A+B+N-1)\ \ldots.\ (A+B)}$$

Etant donné que la deuxième partie du membre de droite de l'équation (40) met en jeu un produit de rapports, B/(A+B), (B+1) / (A+B+1), ... etc, qui sont tous inférieurs à l'unité, R_N tend vers l'unité lorsque le nombre de parutions s'accroît indéfiniment. Ceci implique que la couverture nette estimée par le modèle Bêta Binomial sera toujours inférieure à l'unité, quel que soit le nombre de parutions.

De plus, la moyenne et la variance de la distribution des occasions de voir peuvent être calculées directement à partir des paramètres A, B et N. La moyenne de la distribution des occasions de voir est égale au nombre total d'expositions offertes à la population cible, que l'on désigne sous le nom d'impressions brutes (34).

$$(41) \qquad \begin{matrix}\text{Impressions Brutes} \\ \text{après N partitions}\end{matrix} = GI_N = \Sigma_{K=0}^N \ K*E_K^N = \frac{NA}{A+B}$$

La variance de la distribution des occasions de voir pour N parutions est :

$$(42) \qquad VK_N = \frac{NAB\ (N+A+B)}{(A+B+1)(A+B)^2}$$

Cette expression a des conséquences intéressantes. Supposons, à titre d'illustration, qu'un annonceur ait déterminé que trois expositions sont suffisantes, moins de trois expositions sont inefficaces et plus de trois sont nocives. Supposons, de plus, qu'il ait sélectionné trois supports, chacun d'eux produisant en moyenne trois expositions après six parutions. Quelles sont les valeurs de A et B qui minimisent la variance autour de la moyenne de trois expositions ? La réponse est apportée par l'expression (42), il faut choisir le support possédant les plus fortes valeurs de A et B, comme le montre la table suivante.

Supports	Paramètres A	Paramètres B	Nb. Moyen d'expositions (GI_6)	Variance	Forme de la Distribution
1	0,5	0,5	3	5,25	en U
2	1,0	1,0	3	4,66	plate
3	2,0	2,0	3	3,00	en cloche

Nous pouvons aussi tirer de l'expression (39) la valeur de la fréquence moyenne d'exposition parmi les individus exposés à, au moins, une parution parmi N.

$$(43) \qquad F_N = \frac{GI_N}{R_N} = \frac{NA\ (A+B+N-1)\ldots\ldots\ldots(A+B+1)}{\big((A+B+N-1)\ldots\ldots(A+B)\big) - \big((B+N-1)\ldots\ldots(B)\big)}$$

Cette expression nous permet de répondre à la question suivante : quel est le support qui procure la fréquence moyenne d'exposition la plus élevée parmi les individus exposés au moins une fois ? La réponse est qu'il faut choisir parmi les supports ayant la même audience pour une parution, celui qui a les plus petites valeurs des paramètres A et B. De plus, on observe que, pour une valeur de A donnée, plus le paramètre B est petit, plus la fréquence moyenne est élevée. Par contre, si la valeur de B est donnée, plus la valeur de A est élevée, plus la fréquence moyenne est forte. La table suivante illustre ces résultats, pour des valeurs des paramètres A et B comprises entre 0,2 et 0,7, dans le cas de quatre parutions.

Fréquence Moyenne d'Exposition à Quatre Parutions

Paramètre B	Paramètre A					
	0,2	0,3	0,4	0,5	0,6	0,7
0,2	3,17					
0,3	2,91	2,98				
0,4	2,71	2,78	2,85			
0,5	2,55	2,62	2,69	2,75		
0,6	2,42	2,49	2,56	2,62	2,68	
0,7	2,31	2,37	2,44	2,51	2,56	2,76

Jusqu'ici, nous avons présenté des formules permettant de calculer la couverture nette et la distribution des occasions de voir directement à partir des paramètres A et B de la distribution Bêta binomiale. Nous allons présenter maintenant des équations de récurrence permettant de calculer la couverture nette et la distribution des occasions de voir pour N parutions en fonction de la couverture nette et de la distribution des occasions de voir pour N-1 parutions. Nous commencerons par la distribution des audiences fidèles. A partir de l'équation (39), nous obtenons la relation suivante :

$$(44) \qquad E_{N+1}^{N+1} = \frac{A+N}{A+B+N} E_N^N$$

Il apparaît donc que les audiences fidèles de N+1 parutions sont une fraction des audiences fidèles de N parutions. Puisque le rapport (A+N)/(A+B+N) est toujours inférieur à l'unité, la distribution des audiences fidèles prédite par le modèle Bêta Binomial est continuellement décroissante, ce qui est parfaitement conforme à la logique même du concept d'audience fidèle. La décroissance de l'audience fidèle, d'une parution à une autre, diminue à chaque nouvelle parution ainsi que l'exprime la relation suivante :

$$(45) \qquad E_N^N - E_{N+1}^{N+1} = \frac{B}{A+B+N} E_N^N$$

Ainsi, toutes choses égales par ailleurs, plus le paramètre B est grand, plus la perte en lecteurs fidèles d'une parution à une autre sera élevée.

En ce qui concerne les non lecteurs, chaque nouvelle parution diminuera leur nombre comme suit :

$$E_0^{N+1} = \frac{B+N}{A+B+N} E_0^N \tag{46}$$

Le rapport (B+N)/(A+B+N) étant toujours inférieur à l'unité, le modèle Bêta Binomial prévoit que le nombre de non auditeurs va en diminuant d'une parution à une autre. Ceci est raisonnable puisque les données empiriques montrent que les audiences s'accumulent au cours du temps. Cependant cette accumulation s'arrête bien avant que la population totale soit entièrement exposée, quel que soit le nombre de parutions considérées. Malheureusement, il n'en va pas de même avec le modèle Bêta Binomial qui prévoit qu'éventuellement toute la population soit exposée lorsque le nombre de parutions tend vers l'infini. Il en est ainsi pour la raison que le modèle Bêta Binomial ne prend pas en considération les non lecteurs absolus. Ceci est certainement un point faible du modèle et sur cette base nous pouvons nous attendre à ce que le modèle surestime la couverture nette, spécialement lorsqu'un grand nombre de parutions sera considéré.

Tout comme pour les auditeurs fidèles, nous pouvons calculer la diminution du nombre de non lecteurs consécutive à la N+1 ième parution :

$$E_0^N - E_0^{N+1} = \frac{A}{A+B+N} E_0^N \tag{47}$$

Ainsi, toutes choses égales par ailleurs, plus le paramètre A est grand, plus l'apport de lecteurs nouveaux sera élevé. En d'autres termes, le paramètre A mesure non seulement le pouvoir de rétention des lecteurs fidèles comme l'exprime l'équation (44), mais aussi le pouvoir d'attraction de lecteurs nouveaux comme le montre l'équation (47). La table suivante résume notre analyse concernant la signification des paramètres A et B.

Signification des Paramètres du Modèle Bêta Binomial

Source de Variation	Conséquences	Signification du paramètre
A croît tandis que B reste constant	croissance du % d'auditeurs fidèles et décroissance du % de non auditeurs	Le paramètre A mesure le pouvoir d'attraction et de rétention des auditeurs
B croît tandis que A reste constant	Décroissance du % d'auditeurs fidèles et croissance du % de non auditeurs	Le paramètre B mesure le pouvoir de répulsion et de refoulement des auditeurs

La dernière équation de récurrence présentée nous permet de calculer la couverture nette cumulée de N parutions en fonction de la couverture nette cumulée de N-1 parutions et N-2 parutions

$$R_N = R_{N-1} + \frac{A}{A+B+N-1} \quad E_0^{N-1}$$

$$R_N = R_{N-1} + \frac{B+N-2}{A+B+N-1} \quad (R_{N-1} - R_{N-2}) \tag{48}$$

Il est aussi intéressant de noter que l'on peut calculer par récurrence tous les termes de la distribution des occasions de voir à partir de l'un quelconque d'entre eux et des paramètres A et B. Il suffit pour le démontrer de calculer le rapport de deux termes successifs dans l'équation (39).
Nous obtenons ainsi :

$$E_K^N = \left(\frac{N-K+1}{K} \right) \left(\frac{A+K-1}{B+N-K} \right) E_{K-1}^N \tag{49}$$

Pour clore cette analyse des propriétés mathématiques du modèle Bêta Binomial, nous allons examiner la question subtile de l'apprentissage apparent (35).

- La proportion d'individus qui seront exposés à la N+1 ième parution, parmi ceux qui ont été exposés à toutes les parutions précédentes, est donné par :

$$E_{N+1}^{N+1} / E_{N}^{N} = (A+N) / (A+B+N)$$

- La proportion d'individus qui ne seront pas exposés à la N+1 ième parution, parmi ceux qui n'ont pas été exposés à toutes les parutions précédentes est donnée par :

$$E_{0}^{N+1} / E_{0}^{N} = (B+N) / (A+B+N)$$

Ces deux probabilités conditionnelles croissent avec le nombre de parutions, reflétant ainsi un apprentissage apparent de la part des individus. Il semble que plus on est exposé à un support, plus on a de chances d'y être exposé à nouveau et inversement. Cependant, il ne peut pas en être ainsi dans le cadre du modèle Bêta Binomial puisqu'il a justement été construit sur la base de l'hypothèse de stationnarité des probabilités individuelles d'exposition. La clé de ce paradoxe tient au fait que l'exposition aux parutions successives constitue un processus d'échantillonage séquentiel. A mesure que le nombre de parutions augmente, les individus constamment exposés sont, d'une part, de moins en moins nombreux et, d'autre part, de plus en plus certainement des individus ayant une probabilité d'exposition élevée. Cela ne signifie pas que leurs probabilités d'exposition aient changé. Si nous considérons une seule parution, les individus exposés varieront sans doute largement en ce qui concerne leurs probabilités personnelles d'exposition, en vertu de l'hypothèse d'hétérogénéité. Par contre, lorsque nous considérons N parutions successives, les individus qui ont été exposés à toutes les parutions sont vraisemblablement ceux qui possèdent les probabilités personnelles d'exposition les plus élevées. De la même manière, les individus qui restent non exposés après N parutions possèdent vraisemblablement une probabilité d'exposition faible et, en conséquence, resteront dans doute non exposés à la parution suivante.

L'exactitude de cet argument peut être prouvée en démontrant que le coefficient de corrélation entre les audiences de deux parutions successives reste constant. Formons pour cela la matrice de transition d'une parution à une autre. La première ligne classe les individus non exposés à la N ième parution selon qu'ils ont été exposés ou non à la N-1 ième parution. La seconde ligne fait de même pour les individus exposés à la N ième parution.

autant que du nombre et de l'espacement des répétitions. En résumé, on a découvert que le souvenir est bien davantage affecté par la répétition que ne le sont les comportements, les intentions d'achat ou le nombre de coupons retournés ; que les produits de première utilité (convenience goods) sont plus sensibles aux effets de la répétition que les produits plus secondaires (shopping goods) ; que les réclames agressives agissent mieux sur le souvenir que sur les intentions d'achat, et que les réclames familières utilisant des arguments feutrés obtiennent de meilleures performances, en étant répétées, que les réclames qui surprennent ou qui utilisent des arguments de vente agressifs. Le problème de savoir si les plans d'insertions visant à la saturation conduisent à une usure accélérée du message ou même à des réactions négatives de la part des consommateurs a été examiné, principalement dans le domaine de la radiodiffusion, par Gallup et Robinson (43), Grass (44) et Grass et Wallace (45). L'état des connaissances dans ce domaine a été récemment résumé par Greenberg et Suttoni (46).

Toutes ces études mettent en évidence la nécessité d'inclure des pondérations liées à la fréquence d'exposition ou des fonctions de réponse dans les modèles de planification publicitaire. Cependant, ceci n'est possible que si la partie du modèle consacrée à l'exposition aux supports fournit la distribution des occasions de voir qui est le point de départ nécessaire pour l'application des fonctions de réponse. Ce fait souligne à nouveau l'importance vitale des modèles d'exposition aux supports qui sont étudiés dans cette thèse.

Dernier point, mais non des moindres, un modèle satisfaisant se doit d'inclure des facteurs concurrentiels pour tenir compte des plans d'insertions des concurrents. Un grand nombre d'articles, publiés dans la littérature spécialisée, mettent l'accent sur la nécessité de prendre en compte la réminiscence spontanée (share of mind) (47) réalisée par un plan publicitaire particulier, mais en fait, il n'existe à ce jour qu'une seule étude empirique sur le terrain (48) et qu'une seule étude expérimentale (49) ayant mesuré l'effet de produit par les messages concurrentiels. De même, il n'existe présentement que deux tentatives de modélisation des facteurs concurrentiels (50), (51), (52).

Finalement, outre leur famille, leur philosophie, leur technique et leur portée, les modèles de planification des media diffèrent en termes de critères d'évaluation. La plupart d'entre eux évaluent ou optimisent le choix de la chronologie des insertions sur la base d'un critère unique, alors qu'une approche à critères multiples serait plus appropriée.

Les critères les plus utilisés sont, dans l'ordre décroissant de popularité: l'exposition au support, l'exposition pondérée, la connaissance, le potentiel des ventes, l'essai du produit et le changement d'attitude. Il faut reconnaître qu'aucun d'entre eux, pris séparément, n'est satisfaisant. Ce ne sont que des variables par procuration pour une maximisation du profit. Cependant, on peut faire remarquer que très peu de firmes, sauf dans des situations spécifiques telles que la vente par correspondance de produits inédits, peuvent rattacher, directement et sans ambiguité, leurs ventes et leur profit aux caractéristiques d'un choix particulier de la chronologie des insertions d'un message publicitaire. En conséquence, la plupart des modèles de média laissent la décision budgétaire en dehors du domaine du modèle. De plus, puisqu'il manque à la plupart des modèles les sous-programmes nécessaires d'évaluation du message, ils laissent aussi à l'écart les problèmes de sélection de message. Cet état de l'art est des plus malheureux, vu qu'il est largement reconnu que les média et les facteurs de message interagissent (40), (41), (42). Cependant, même les critères d'efficacité de communication formulés en termes de connaissance sont considérés par beaucoup comme trop ambitieux, et ils se contentent de maximiser l'exposition du message publicitaire. En fait, même cette tâche limitée apparaît formidable, vu la pauvreté des informations concernant l'exposition au message. En conséquence, la plupart des modèles de media en viennent à maximiser l'exposition au support, comme succédané de meilleures variables. Cependant, cela est déjà plus satisfaisant que de minimiser le coût par millier d'insertions, du moment qu'on utilise un modèle d'exposition aux média, dûment validé.

Il faut noter que le choix de la fonction objective n'est pas indépendant de la portée du modèle. Il est évident qu'un modèle qui incorpore seulement l'audience et les facteurs d'exposition ne peut pas, quelle que soit l'ingéniosité de son constructeur être très utile pour prédire la réaction des ventes. Il faut aussi se rendre compte que plus la portée du modèle est vaste, plus les données nécessaires à son fonctionnement seront sujettes à caution. En conséquence, il faut réaliser une sorte d'équilibre entre la portée du modèle et la quantité d'opinions subjectives que l'utilisateur du modèle est capable ou désireux d'incorporer au modèle. Cette remarque présentée en premier lieu par Kotler (53), a été élaborée par Broadbent et Segnit (54) et par Rothman (55).

On peut trouver un bon exemple d'un tel équilibre dans le concept de "couverture nette efficace" utilisé à l'agence J.W.T. (56). La couverture nette réelle d'un plan de campagne publicitaire comprend seulement les individus-cibles exposés suffisamment souvent pour assurer une réaction potentielle. Elle exclut

les individus insuffisamment exposés et les individus surexposés.

Nous terminons notre discussion en présentant les caractéristiques souhaitables qu'un modèle de média planning idéal se devrait de posséder. Selon Little (57), un bon modèle de média devrait être facile à comprendre, robuste, de contrôle aisé et adaptable à de nouvelles sources de données. Il devrait aussi être complet, ce qui veut dire qu'il devrait laisser à l'utilisateur le choix quant aux jugements subjectifs devant être insérés le cas échéant. De plus, il devrait être aisé de communiquer avec le modèle par l'intermédiaire d'un terminal acceptant des commandes formulées en Français.

A ceci, il convient d'ajouter qu'un bon modèle de media devrait être minutieusement validé. Compte tenu de l'ensemble de ces facteurs, nous sommes en mesure d'examiner les principaux modèles de media planning de ces 15 dernières années.

13. Tour d'horizon des principaux modèles de planification publicitaire.

Nous avons choisi 14 modèles parmi les plus proéminents. Ils sont présentés par ordre chronologique dans la table 14. La première partie de la table compare les modèles en terme de famille, de philosophie, de technique, de mise en oeuvre informatique, de facteurs d'audience et de facteurs d'exposition. La seconde partie de la table compare les modèles en termes de facteurs de message, de facteurs de fréquence, de facteurs compétitifs et de critères d'évaluation.

Les progrès accomplis au cours des 15 dernières années ont pris forme dans les directions suivantes :

- davantage de souplesse dans la définition des segments-cibles, tant sur le plan socio-démographique que sur le plan géographique ou sur le plan de la consommation du produit ;

- davantage de profondeur psychologique aussi bien dans les données que dans les modèles ;

- une portée accrue des modèles, spécialement dans les domaines des facteurs d'audience et des facteurs d'exposition ;

- de l'optimisation rigide de modèles incomplets vers des méthodes heuristiques

flexibles appliquées à des modèles plus réalistes ;

- du "batch processing" vers des systèmes conversationnels et interactifs ;

- de larges programmes tous usages vers des bibliothèques de sous-programmes spécifiques, mieux adaptés à la variété des situations marketing (58) ;

- un rôle plus actif de l'utilisateur par le biais du contrôle des paramètres adaptatifs du modèle.

Comme le montre la table 14, il reste encore beaucoup à faire. Dans les dix années à venir, les améliorations futures porteront vraisemblablement sur les thèmes suivants :

- des données d'audience plus récentes et de meilleure qualité ;
- de meilleurs modèles de prévision d'audience (59) ;
- une amélioration des comparaisons inter-media (60) ;
- une amélioration des données portant sur l'exposition au message (61) ;
- une amélioration des mesures de la perception des messages (62) ;
- des fonctions de réponse plus précises et plus différenciées selon les media, les catégories de produits et les critères d'efficacité, remplaceront les fonctions de réponses globales et unicritères ;
- un passage de l'évaluation unicritère à l'évaluation multicritère des plans publicitaires.

Nous avons maintenant terminé notre discussion de l'état de l'art en matière de planification des media. Nous avons présenté les différentes complexités du problème et les différentes solutions proposées pour les résoudre. Ce faisant, nous avons dévoilé le rôle essentiel joué par le concept d'exposition dans les modèles de planification publicitaire. Nous allons maintenant analyser les divers modèles d'exposition aux media.

TABLE 14

TOUR D'HORIZON DES PRINCIPAUX MODELES DE PLANIFICATION PUBLICITAIRE

Modèle	Famille	Philosophie	Technique	Utilisation ordinateur	Facteurs d'audience	Facteurs d'exposition
I.P.A. 1958	Media Selection	Evalue	Tabulation	Batch Processing	Demographiq Geographiq	Non
LP-1 1961	Media Selection	Optimise	Linear Programm.	Batch Processing	Demographiq Geographiq	Pondérations d'exposition
BEA 1961-63	Media Selection	Optimise	Heuristic	Nomogramm. (non computer	Demographiq	Véhicules et Ad Expos.pond.
HIGH ASSAY 1961	Media Selection	Optimise	Heuristic	Batch Processing	Demographiq Geographiq	Véhicules Pondérations
C.A.M. 1964	Media Planning	Evalue	Simulation	Batch Processing	Demographiq Product Usage	Véhicules et Ad Exp.pond.
Brown & Warshaw-65	Media Select.	Optimise	Non linear Program	Batch Processing	Demographiq Ventes potentiel	Pondérations d'expos.
SCAL 1967	Media Planning	Evalue Optimise	Simulation	Batch Processing	Demographiq Ventes potentiel	Véhicules Pondérat.
MARKET MATH-67	Media Selection	Evalue	Trial Schedules	Batch Processing	Demographiq Ventes potent.	Véhicules et Ad Exp.pond.
LP-2 1968	Media Planning	Optimise	Goal Program	Batch Processing	Demographiq Ventes potent.	Véhicules Pondérat.
MEDIAC 1969	Media Planning	Optimise	Heuristic Program	On Line Time Share	Demographiq Geographiq Ventes potent.	Véhicules et Ad Exp.Pond.
AD.ME.SIM 1969	Media Planning	Evalue	Simulation	Batch Processing	Demographiq Ventes Potent.	Véhicules et Ad Exp.Pond
COSMIC 1970	Media Selection	Evalue Optimise	Heuristic Program	Batch Processing	Demographiq Ventes Potent.	Véhicules Pondérat.
I.M.S. 1971	Media Selection	Evalue Optimise	Branch & Bound Program	On Line Time Share	Demographiq Psychographiq Geographiq Ventes Potent.	Véhicules Pondérat. Ad Expo.Pond.
ADMOD 1975	Media Sélection	Optimise	Heuristic Program		Trial Value Weight	Vehicules et Ad Expos. Pond.

TABLE 14
(suite)

TOUR D'HORIZON DES PRINCIPAUX MODELES DE PLANIFICATION PUBLICITAIRE

Media	Facteurs de media	Facteurs Message	Facteurs Fréquence	Facteurs Competition	Critères d'évaluation
I.P.A. 1958	Non	Non	Non	Non	Coût aux 1000 Lecteurs utiles
LP-1	Pondérat. de media	Non	Non	Non	Coût pondéré par lecteur utile
BEA 1961-63	Non	Format.	Réponse Géométrique	Non	Couverture ou Fréquence Moy.
HIGH ASSAY 1961	Pondérat. de media	Format.	Non	Non	Couverture ou Fréquence Moy.
C.A.M. 1964	Pondérat. de media	Attention	Réponse géométrique	Non	Indice d'efficacité
Brown & Warhaw-65	Pondérat. de media	Non	Non	Non	Coût pondéré par lecteur utile
SCAL 1967	Non	Non	Non	Non	Multi critères d' évaluation
MARKET-MATH-67	Non	Non	Pondération de fréquence	Non	Couverture et distribut. de fréq.
LP-2 1968	Pondérat. de media	Non	Fréquence Moyenne	Non	Pondérat. mixte de couverture et fréq.
MEDIAC 1969	Pondérat. de media	Non	Courbe de réponse	Non	Ventes potentielles
AD.ME.SIM 1969	Pondérat. de media	Format. Position	Courbe de réponse	Non	Indice d'impact
COSMIC 1970	Pondérat. de media	Non	Courbe de réponse	Non	Indice d'impact
I.M.S. 1971	Pondérat. de media	Format et position	6 différentes courbes de réponse	Non	Exposition pondéré ou indice d'impact
ADMOD 1975	Non	Non	Courbe de réponse	Non	Indice d'incitation à l'achat

2. Les modèles d'exposition aux media.

21. Le modèle implicite.

Certains auteurs se contentent d'assimiler l'insertion avec l'exposition. En d'autre termes, la nécessité de disposer d'un modèle d'exposition n'est même pas reconnue. Le raisonnement est le suivant. Si le plan publicitaire prévoit une insertion dans le magazine Newsweek, et si l'audience moyenne d'une parution de Newsweek est de 20 millions de lecteurs, alors 20 millions d'expositions auront été réalisées. Ce raisonnement n'est valable que tant que le plan publicitaire ne comprend qu'une seule insertion dans un seul support, ce qui est rarement le cas. Dès que l'on considère deux insertions, soit dans le même support, soit dans deux supports différents, il n'en va plus ainsi et le nombre d'expositions n'est pas égal à deux fois l'audience du support ni à la somme des audiences des deux supports. Ceci résulte du fait qu'un grand nombre de lecteurs (ou d'auditeurs) lisent (ou regardent) les deux parutions ou les deux supports. Le premier phénomène, appelé accumulation d'audience, est lié au fait que l'audience d'un support ne se renouvelle jamais dans sa totalité d'une parution à une autre. De même, l'audience ne reste pas totalement identique d'une parution à une autre. La vérité se situe quelque part entre ces deux extrêmes, et ce "quelque part" varie d'un support à un autre et d'une période à une autre. Une définition plus précise de l'accumulation d'audience et des diverses manières de la mesurer sera présentée dans le chapitre 4. Le second phénomène, appelé duplication d'audience, est lié au fait que la plupart des supports ont des lecteurs (ou des auditeurs) communs. L'audience commune à deux supports est appelée audience dupliquée. Il ne s'agira que de quelques individus lorsque les deux supports considérés sont spécialisés chacun dans des domaines fort éloignés. Par contre, des supports spécialisés dans un même domaine obtiendront des audiences quasiment identiques. Par exemple, le Journal of Marketing et le Journal of Greek Philosophy ont vraisemblablement peu de lecteurs communs, alors que la plupart des lecteurs du Journal of Marketing lisent aussi le Journal of Marketing Research. Une définition précise de la duplication d'audience sera présentée dans le chapitre 4. Ici, il suffit de noter que lorsqu'on rend égales l'insertion et l'exposition, on assume implicitement les deux hypothèses suivantes :

- Les audiences de chaque support se renouvellent entièrement d'une parution à une autre ;

- Chaque support possède une audience exclusive qu'il ne partage avec

aucun autre support.

Ces deux hypothèses sont de toute évidence absolument irréalistes et conduisent à une tragique surestimation du nombre des individus exposés aux divers supports utilisés par la campagne publicitaire. Dans le même temps, le fait que certains individus soient exposés plus d'une fois est absolument négligé. Pourtant, plusieurs modèles de planification publicitaires sont tombés dans ce piège (63),(64), (65), (66), (67), (68), (69), (70). Néanmoins, l'honnêteté envers ces auteurs oblige à reconnaître que la plupart d'entre eux étaient parfaitement conscients des inconvénients du modèle implicite d'exposition. Ils l'utilisèrent néanmoins à cause de l'hypothèse de linéarité de la technique d'optimisation (programmation linéaire) dont ils s'étaient fait les avocats.

Kotler (53) fut l'un des premiers à dénoncer cette erreur qui devint par la suite un exemple classique de la tendance à asservir le problème à résoudre à la technique disponible plutôt que de faire le contraire.

Zangwill (71) montra comment la programmation décisionnelle peut être utilisée pour tenir compte explicitement des duplications d'audience. Cependant, son approche ne tient pas compte de l'existence des triplications et, d'une manière générale, des recouvrements d'audience d'ordre supérieur à deux qui apparaissent lorsque l'on considère plus de deux supports. A cause de cette faiblesse théorique, liée à l'incapacité de la programmation décisionnelle à manipuler des plans d'insertions importants, son modèle n'atteignit pas le stade des applications.

L'existence des phénomènes de duplication et d'accumulation d'audience étant explicitement reconnue, trois catégories de modèles ont été proposées pour en tenir compte.

22. Le modèle naïf.

L'hypothèse la plus simple qui peut être faite au sujet de l'accumulation et de la duplication des audiences est de supposer qu'elles se produisent au hasard. En d'autres termes, le modèle naïf suppose que la proportion des lecteurs ou des auditeurs d'une deuxième parution ou d'un second support est la même parmi les lecteurs de la première parution que parmi les non-lecteurs. Ainsi, selon ce modèle, les audiences de divers supports, de même que les audiences de diverses parutions d'un même support, sont indépendantes au sens statistique du terme.

L'exposition est réduite à un évènement parfaitement aléatoire.

Malheureusement, toutes les données disponibles pointent dans une autre direction. Ce point, qui sera établit en détail aux chapître 5, 6 et 7, est connu depuis le début des années 50, avec la publication des enquêtes de Politz sur l'audience du magazine LIFE (72). Depuis cette date, de très nombreuses études d'audience effectuée en Angleterre, en Allemagne, en France, aux USA, au Canada, au Japon et en Australie ont fourni des résultats identiques qu'il serait fastidieux d'énumérer ici.

Il est facile d'expliquer pourquoi les audiences ne s'accumulent pas au hasard. La fidélité des lecteurs ou des auditeurs à leur support préféré nous fournit la réponse. Tout d'abord, de nombreuses publications sont vendues par abonnement et, même celles qui sont achetées dans des kiosques, s'attirent des lecteurs fidèles, soit qu'elles constituent l'unique source disponible pour certains types d'informations, soit qu'elles bénéficient du penchant des hommes à se créer des habitudes. La fidélité des audiences ne se limite pas aux supports imprimés, elle se vérifie aussi, à des degrés divers, pour les spectacles télévisés, spécialement les feuilletons et les émissions d'information. La même observation s'applique aussi aux stations de radio qui développent la fidélité de leurs audiences par la sélection judicieuse d'un format musical, d'un dosage d'informations et de divertissements et, enfin et surtout, par une promotion permanente de leurs présentateurs. Même la publicité extérieure bénéficie d'audiences fidèles puisque les personnes qui se déplacent biquotiennement pour leur travail empruntent toujours le même trajet.

Jusqu'ici nous n'avons traité que de l'accumulation, mais des observations similaires s'appliquent à la duplication des audiences. La première étude approfondie de la duplication, elle aussi due à Politz (73) a été suivie d'enquêtes identiques menées par l'I.P.A. en Angleterre (74) et par le CESP en France (75). Ces enquêtes ont montré qu'il existe une corrélation positive entre les audiences des supports orientés vers des centres d'intérêts similaires et une corrélation négative entre les audiences des supports axés sur des centres d'intérêts sans rapport. Cette observation a été vérifiée systématiquement dans toutes les enquêtes annuelles d'audience réalisées par les bureaux de recherches de Simmons et de TGI. La tendance observée au cours des quinze dernières années indique une accentuation du phénomène. Cette accentuation se traduit par la disparition progressive des magazines "grand public" tel que LIFE ou Saturday Evening Post. Si ce mouvement se maintient, le concept de communication de masse

sera bientôt vidée de son sens dans les dix années à venir, exception faite des audiences télévisées. Deux facteurs peuvent être avancés pour expliquer la prédominance des corrélations positives entre les audiences. Le premier facteur est l'existence dans la population de centres d'intérêts tel que le sport, la mode, les informations, le sexe, etc... Ces centres sont activement recherchés et développés par les éditeurs car ils signifient d'une part audiences fidèles et d'autre part des revenus publicitaires en provenance des industriels qui écoulent les produits qui répondent à ces besoins particuliers. Le second facteur est l'existence de diverses propensitées à la lecture (ou à l'écoute) dans les diverses couches de la population. Un individu qui lit (ou écoute) beaucoup, ne se cantonne pas en général dans un domaine d'intérêt unique. Par contre, un individu qui lit (ou écoute) peu, se cantonne dans son domaine de prédilection.

Cependant, malgré les montagnes de données accumulées, (76), (77), (78), (79), (80), (81), (82), (83), le modèle de l'exposition aléatoire conserve des partisans (84), (85), (86), (87), (88), (89), qui font valoir la simplicité du modèle et la facilité des calculs qui en résulte. Toutefois, par son refus sans fondement des habitudes de lecture (ou d'écoute), le modèle naïf d'exposition s'est attiré les foudres des praticiens (90). Cependant, un des défenseurs du modèle demeure apparemmant inébranlé (91).

En résumé, il apparaît que les individus ne se partagent pas nettement en lecteurs et non lecteurs comme le modèle implicite le laisse entendre, et qu'ils ne se conduisent par hasard comme le modèle naïf voudrait nous le faire croire. En conséquence, une troisième catégorie de modèles d'exposition est apparue, faisant appel à une évaluation subjective des habitudes de lecture (ou d'écoute).

23. Le modèle subjectif.

Plutôt que de faire appel à des hypothèses simplificatrices toujours appauvrissantes, les tenants de l'approche subjective insistent sur la nécessité de s'appuyer sur les appréciations des experts chevronnés. Des pondérations sont utilisées pour escompter l'audience cumulée à partir de l'audience brute. Ces pondérations reflètent l'expérience accumulée par le spécialiste de la planification des supports. Chaque agence de publicité a construit ainsi, au cours du temps, de nombreuses tables indiquant, support par support, le pourcentage de personnes exposées et la fréquence moyenne d'exposition résultant de N parutions.

Kotler (53) fut le premier à soutenir cette approche, argumentant d'une manière convaincante qu'en l'absence de données solides, le recours à des estimations subjectives est préférable à une confiance aveugle en des modèles contestables. En d'autres termes, même si les pondérations subjectives sont imprécises, elles permettront toutefois d'arriver à de meilleures décisions qu'en ignorant tout bonnement les accumulations et les duplications ou qu'en les réduisant à des phénomènes purement aléatoires. Cependant, l'approche subjective n'est pas sans problèmes. Le plus souvent, les pondérations d'expositions ont été confondues avec les pondérations de fréquence (92). Par exemple, le fait de donner une pondération moindre à la seconde insertion qu'à la première reflète-t-il le fait que la seconde parution apporte moins de lecteurs nouveaux que la première, ou reflète-t-il le fait qu'un lecteur exposé deux fois a moins d'importance que deux lecteurs exposés chacun une fois ? En toute logique, ces deux aspects sont différents et doivent être distingués. Le premier porte sur une question de faits, combien de personnes seront exposées, combien de fois, alors que le second porte sur la valeur attribuée par l'annonceur aux expositions successives.

Quoique le recours aux pondérations subjectives soit une pratique courante permettant d'intégrer à la fois les facteurs d'audience, d'exposition, de message et de fréquence, il faut bien reconnaître que ces pondérations ne sont qu'un pis aller traduisant l'insuffisance de nos connaissances. Selon l'expression de Broadbent et Segnit (93), elles ne sont que "la poubelle dans laquelle on déverse tout ce qui n'a pu être explicitement inclus dans le modèle". Les modèles d'expositions implicites et naïfs étant incontestablement faux, il a bien fallu les amender par l'introduction de pondérations subjectives. Néanmoins, les pondérations ne constituent pas en elle même un authentique modèle d'exposition aux supports publicitaires. Leur rejet devint inévitable avec l'arrivée de meilleurs modèles d'exposition.

24. Les modèles sophistiqués.

Trois classes de modèles d'exposition appartiennent à cette catégorie, les modèles de type Ad Hoc, les modèles stochastiques et les modèles de simulation. Tous ces modèles sont sophistiqués au sens suivant. Ils admettent l'existence des accumulations et des duplications d'audience et reconnaissent qu'elles ne sont pas le fruit du pur hasard. Ils utilisent tous des données qui vont au delà de l'audience moyenne d'une parution dans la détermination des comportements d'exposition. Toutefois, la manière dont ces données sont traitées varie d'une classe de modèle à une autre.

Les modèles de type Ad Hoc se contentent d'ajuster des fonctions simples aux données disponibles dans le but d'extrapoler. Les modèles stochastiques essayent de réduire les comportements d'exposition à des lois de probabilité connues, sur la base d'hypothèses simplificatrices. Les modèles de simulation tentent d'affecter des probabilités personnelles d'exposition à un échantillon d'individu afin de reproduire la réalité des comportements d'exposition sous la forme d'un microcosme manipulable. Les premiers modèles d'exposition des années 60 appartenaient en majorité à la première catégorie. Ils traitaient, soit des duplications (76), (94), soit des accumulations (77), (84). Le premier modèle stochastique de duplication fut présenté en 1963 (95). Le premier modèle stochastique d'exposition traitant simultanément l'accumulation et la duplication apparut en 1964 (96) et fut suivi en 1966 par le premier modèle complet d'exposition du type Ad Hoc (97). L'utilisation des méthodes de simulation fut préconisé dès 1961 par Simulmatic Corporation (98), (99), mais dû attendre 1966 pour devenir parfaitement opérationnelle (88). Plusieurs autres modèles de simulation apparurent ensuite (100), (101), (102), (103), (104), (105), caractérisés par une absence quasi générale d'efforts de validation empirique, à l'exception des modèles CAM (88) et SCAL (106) et du modèle de simulation des probabilités personnelles (90).

Les modèles les plus récents tendent à abandonner les méthodes de simulation au profit de modèles stochastiques de plus en plus complexes(107), (108), (109), (110), (111), (112), (113), (114), (115), (116), (117), (118). Tous les modèles d'exposition que nous venons de présenter brièvement sont étudiés en détail dans les chapitres 5, 6 et 7 de cette thèse. Leurs hypothèses, ainsi que leurs performances empiriques, sont examinées et évaluées. Cette étude comparative est la première du genre puisqu'aucun de ces modèles d'exposition n'avaient auparavant été testé et confronté sur la base de banques de données homogènes. Précisont, toutefois, que notre étude ne s'étend pas aux modèles de planification des supports dans leur totalité mais uniquement à la partie de ces modèles consacrée à l'exposition aux supports. On peut trouver une excellente comparaison des mérites théoriques des principaux modèles de planification des media chez Gensch (119). C'est précisément l'absence de toute étude empirique d'ensemble qui a motivé notre recherche.

Cependant, les objectifs de notre étude vont au delà de la simple comparaison des performances empiriques des divers modèles d'exposition. Ils sont exposés en détail dans le chapitre suivant.

NOTES

(1) - BROADBENT S.R. "Media planning and computers by 1970" The Thomson Medals and Awards for Media Research, London, 1965, pp. 61-86.

(2) - GENSCH D.H. "Computer models in advertising media selection", Journal of Marketing Research, vol. 5, Nov. 1968, pp. 414-424.

(3) - GENSCH D.H. "Media factors, a review article", Journal of Marketing Research, vol. 7, Mai 1970, pp. 216-225.

(4) - BROADBENT S.R. "Beyond cost per thousand - an examination of media weights", The Thomson Medals and Awards for Media Research, London, 1968, pp. 105-141.

(5) Interactive Market Systems Inc, 19 West 44 th Street, New York.

(6) - Telmar Communication Corporation, 90 Park Avenue, New York.

(7) - Centre d'Information sur les Media, dossier d'information n° 3, Bruxelles, SOBEMAP, Janvier 1975.

(8) - SISSORS J.Z. "Matching Media with Markets", Journal of advertising Research, vol. 11, October 1971, pp. 39-43.

(9) - POLITZ A., Media Studies, Ad Page Exposure in 4 Magazines, 1960, New York.

(10) - SIMMONS W.R. & Associates Research Inc., Selective Markets and the Media Reaching Them, New York, 1969.

(11) - LUCAS D.B., BRITT S.H., Measuring Advertising Effectiveness, Mc Graw-Hill New York, 1963.

(12) - BANKS S. "Patterns of Daytime Viewing Behavior", Proceeding of the A.M.A. Conference, June 1967, pp. 139-142.

(13) - Swanson C.E. "Patterns of Nightime Television Viewing", Proceedings of the A.M.A. Confernence, June 1967, pp. 143-147.

(14) - FRANK R.E., BECKNELL J.C., CLOKEY J.D., "Television Program Types", Journal of Marketing Research, vol. 8, May 1971, pp. 204-211.

(15) - TIGERT D.J., "Are television Audiences Really Different ?", Proceeding of the A.M.A. Conference, April 1971, pp. 239-246.

(16) - WELLS W.D., Banks S., Tigert D.J. "Order in the Data", Proceeding of the A.M.A. Conference, June 1967, pp. 263-267.

(17) - WELLS W.D., "The Rise and the Fall of Television Program Types", Journal of Advertising Research, Vol. 9, Sept. 1969.

(18) - WILSON C.L., "Homemaker Living Patterns and Market Place Behavior : A Psychometric Approach", Proceedings of the A.M.A. Conference, Spring 1966, pp. 305-331.

(19) - BASS F.M., Pessemier E.A., Tigert D.J. "A Taxonomy of Magazine Readership applied to problems in Marketing Strategy and Media Selection", Journal of Business, Vol. 42, July 1969, pp. 337-363.

(20) - BERNAY E.K. "Life Style Analysis as a basis for Media Selection", in King & Tigert eds., Attitude Research Reaches New Heights, 1971, 189-195.

(21) - KING C.W., Summers J.O., "Attitudes and Media Exposures", Journal of Advertising Research, Vol. 11, February 1971, pp. 26-32.

(22) - TIGERT D.J., "A psychographic Profile of Magazines Audiences : An Investigation of e Medium Climate", 1st Annual Conference, Association for Consumer Research, August 1969.

(23) - STARCH D., Measuring Advertising Readership and Results, Mc Graw Hill, 1966.

(24) - DIAMOND D.S., "A Quantitative Approach to Magazine Advertissement Format Selection", Journal of Marketing Research, Vol. 5, Nov. 1968, pp. 376-386.

(25) - SCHONFELD E.P., "Prediction of Readership and Enquiry Response to Print Advertising", Ph. D. Thesis, Nortwestern University, August 1975.

(26) - PENINOU G., Intelligence de la Publicité : Etude Sémiotique, Collection Media et Messages, Robert Laffont, 1972, Paris.

(27) - BOGART L. Ed., Psychology in Media Strategy, American Marketing Association, 1966, Chicago.

(28) - WRIGHT P.L., "On the Direct Monitoring of Cognitive Response to Advertising", in Buyer/Consumer Information Processing, Hughes & Ray eds., University of North Carolina Press, 1974.

(29) - MALONEY J.C., "Is Advertising Believability Really Important ?", Journal of Marketing, Vol. 27, Octobre 1963, pp. 1-8.

(30) - MALONEY J.C., "Copy Testing : What Course is it Taking ?", Proceedings of the 9 th Annual Conference of the A.R.F., New York, 1964.

(31) - AXELROD J.N., Advertising Research Foundation Guide to Copy Research, A.R.F., New York, September 1971.

(32) - GROSS I., "The Creative Aspects of Advertising", Sloan Management Review, Vol. 14, Fall 1972, pp.83-110.

(33) - EBBINGHAUSS H. Uber das Gedachtnis, Duncker and Humbolt, Leipzig, 1885.

(34) - STRONG E.K., "Size and Frequency in Advertising", Association of National Advertising Manager, Research Bull. N° 4, April 1913.

(35) - DURAND J., "Advertising and Time : Growth and Decay", The Marcel Dassault Awards, Paris, 1969, pp. 47-93.

(36) - ZIELSKE H.A., "The Remenbering and Forgetting of Advertising", Journal of Marketing, Vol. 23, January 1959, pp. 240-244.

(37) - HUGHES M., "What Time Interval should Elapse between different Insertions in the same Magazine", The Marcel Dassault Awards, Paris, 1969, 13-42.

(38) - STRONG E.C., "The Use of Field Experimental Observations in Estimating Advertising Recall", Journal of Marketing Research, Vol. 11, November 1974, pp. 369-378.

(39) - DOMARACKI J., "BLAZT, A User's Guide to the Leo Burnett Company Scheduling Model", Internal Memo, Leo Burnett Co, Chicago, 1971.

(40) - RAY M.L., Sawyer A.G., "Repetition in Media Models : A Laboratory Technique", Journal of Marketing Research, Vol. 8, February 1971, pp. 20-29.

(41) - RAY M.L., SAWYER A.G., "Behavioral Measurements for Marketing Models : Estimating the Effects of Advertising Repetition for Media Planning", Management Science, Vol. 18, December 1971, pp. 73-89.

(42) - RAY M.L., SAWYER A.G., STRONG E.C., "Frequency Effects Revisited", Journal of Advertising Research, Vol. 11, February 1971, pp. 14-20.

(43) - GALLUP & ROBINSON, Inc., An Analysis of Television Comercial Build-Up/Wear-Out., Princeton, March, 1967.

(44) - GRASS R.C., "Satiation Effects of Advertising", Proceeding of the 14 th Annual Conference of the A.R.F., New-York, October 1968, pp. 20-28.

(45) - GRASS R.C. WALLACE W.H., "Satiation Effects of Tv Commercials", Journal of Advertising Research, Vol. 9, June 1969, pp. 3-8.

(46) - GREENBERG A., SUTTONI C., "Television Commercials Wear-Out", Journal of Advertising Research, Vol. 13, October 1973, pp. 47-54.

(47) - GRUBER A., "Top of Mind Awarness and Share of Families : an Observation", Journal of Marketing Research, Vol. 7, May 1969, pp. 227-231.

(48) - STARCH D., Reading Competitive Advertissements, Starch Ad Readership Service, N° 58, New-York.

(49) - CHINSKY J.M., WELLS W.D., "Effects of Competing Messages : A Laboratory Simulation", Journal of Marketing Research, May 1965, pp. 141-148/

(50) - LODISH L.M., "Considering Competition in Media Planning", Management Science, Vol. 17, February 1971, B. 293- 306.

(51) - LODISH L.M., "Exposure Interactions among, Media Schedules", Journal of Advertising Research, Vol. 13, April 1973, pp. 31-34.

(52) - CHANDON J.L., "M.T.M. : A Media Timing Model", CESAM Working Paper, Louvain, 1973.

(53) - KOTLER P., "Toward an Explicit Model for Media Selection", Journal of Advertising Research, Vol. 4, March 1964, pp. 34-41.

(54) - BROADBENT S.R., SEGNIT S., "Response Functions in Media Planning", The Thomson Medal and Awards for Media Research, London, 1967, pp. 35-86.

(55) - ROTHMAN L.J., "The Use of Response Functions : A discussion and an Alternative", The Thomson Medal and Awards for Media Research, London 1967, 87-122.

(56) - The Concept of Effective reach, Unauthored internal memo, J.W.T., Nov. 1973.

(57) - Little J.D.C., "Models and Managers : The Concept of a Decision Calculus", Management Science, Vol. 16, Apris 1970, B. 466-485.

(58) - CATRY B., CHEVALIER M., "The Evolution of French Media Models", Journal of Advertising Research, Vol. 13, June 1973, pp. 19-26.

(59) - FEDERMAN S.P., "Filling the Syndicated Data Gap", Proceedings of the 19 th Annual Conference of the A.R.F., New-York, 1974.

(60) - RAMOND C., "Toward Valid Intermedia Comparisons of Response to Advertising", The Marcel Dassault Award, Paris, 1970.

(61) - The T.G.I., Reader Quality Study, T.G.I., New York, 1975.

(62) - HENSEL J., STERNTHAL B., "Physiological Measures of Information Receptivity : Galvanic Skin Response and Pupil Dilatation", Working Paper, Northwestern University, 1974.

(63) - MORAN W.T., "Practical Media Decision and the Computer", Journal of Marketing, Vol. 27, July 1963.

(64) - DAY R.L., "Linear Programming in Media Selection", Journal of Advertising Research, Vol. 2, June 1962, pp. 40-44.

(65) - ENGEL J., WARSCHAW M.R., "Allocating Advertising Dollar by Linear Programming", Journal of Advertising Research, Vol. 4, September 1966, pp. 42-48.

(66) - BROWN D.B., WARSHAW M.R., "Media Selection by Linear Programming", Journal of Marketing Research, February 1965, pp. 83-88.

(67) - STASCH S.F., "Linear Programming and Space-Time Considerations in Media Selection", Journal of Advertising Research, Vol. 15, December 1965, pp. 40-46.

(68) - BASS F.M., LONSDALE R.T., "An Exploration of Linear Programming in Media Selection", Journal of Marketing Research, May 1966, pp. 179-188.

(69) - STASCH S.F., "Linear Programming and Media Selection : A Comment", Journal of Marketing Research, May 1967, pp. 205-207.

(70) - MENKE W.W., "Linear Programming Applied to Marketing Decision Making", Proceedings of the American Marketing Association, June 1968, pp. 245-252.

(71) - ZANGWILL W.I., "Media Selection by Decision Programming", Journal of Advertising Research, Vol. 5, September 1965, pp. 30-36.

(72) - POLITZ Alfred, Research, Inc., A Study of the Household Accumulative Audience of Life, New-York, 1952.

(73) - POLITZ Alfred, Research, Inc., Life Study of Consumer Expenditures, New York, 1957.

(74) - Institute of Practitioners in Advertising, National Readership Survey, London.

(75) - Centre d'Etude des Supports de Publicite, Paris, 1957.

(76) - AGOSTINI J.M., "How to Estimate Unduplicated Audiences", Journal of Advertising Research, Vol. 1, March 1961, pp. 11-14.

(77) - AGOSTINI J.N., "Analysis of Magazine Accumulative Audience", Journal of Advertising Research, Vol. 2, October 1962, pp. 24-27.

(78) - BANKS S., MANDANSKY A., "Estimation of Multimagazine Readership", Journal of Business, July 1958.

(79) - BOWER J., "New Audiences of U.S. and Canadian Magazines : Seven Test of Agostini's Formula", Journal of Advertising Research, Vol. 3, March 1963, pp. 13-20.

(80) - KUHN W., "Net Audiences of German Magazines : A New Formula", Journal of Advertising Research, Vol. 3, March 1963, pp. 30-33.

(81) - KAATZ R.B., "Improving Agostini Formula for New Audience". Letter to the Editor, Journal of Advertising Research, Vol. 3, September 1963.

(82) - MARC M., "Net Audiences of French Business Papers : Agostini's Formula Applied to Special Markets", Journal of Advertising Research, Vol. 3, March 1963, pp. 26-29.

(83) - ENGELMAN, F.L., "An Empirical Formula for Audience Accumulation", Journal of Advertising Research, Vol. 5, June 1965, pp. 21-28.

(84) - LEE A.M., BURKART A.J., "Some Optimization Problems in Advertising Media Planning", Operational Research Quarterly, Vol. 11, 1960, pp. 113-122.

(85) - LEE A.M., "Decision Rules for Media Scheduling : Static Compaigns", Operational Research Quarterly, Vol. 13, 1962, pp. 229-242.

(86) - TAYLOR C.J., "Some Development in the Theory and Application of Media Scheduling Methods", Operational Research Quarterly, Vol. 14, 1963, pp. 291-305.

(87) - ELLIS D.M., "Building up a Sequence of Optimum Media Schedules", Operational Research Quarterly, Vol. 16, 1965, pp. 413-424.

(88) - BEALE E.M.L., HUGHES P.A.B., BROADBENT S.R.M., "A Computer Assessment of Media Schedules", Operational Research Quarterly, Vol. 17, 1966, pp. 381-411.

(89) - AAKER D.A., "A probabilistic Approach to Industrial Media Selection", Journal of Advertising Research, Vol. 8, September 1968, pp. 46-54.

(90) - GREENE J.D., "Personal Media Probabilities", Journal of Advertising Research, Vol. 10, October 1970, pp. 12-18.

(91) - AAKER D.A., "Admod : An Advertising Decision Model", Journal of Marketing Research, vol. 12, Février 1975, pp. 37-45.

(92) - LUCAS D.B., "Mathematical Programming for Better Media Selection", Address to the American Association of Advertising Agencies, November, 1961, New-York.

(93) - BROADBENT S.R., SEGNIT S., "Beyond Cost per Thousand-An Examination of Media Weights", The Thomson Medals and Awards for Advertising Research 1968, pp. 105.

(94) - STARCH Daniel & STAFF, Inc., Starch Magazine report, 1967, New-York.

(95) - CAFFYN J.M., SAGOVSKY M, "Net Audiences of British Newspapers : A Comparison of the Agostini and Sainsbury Methods", Journal of Advertising Research, vol. 3, March 1963, pp. 21-24.

(96) - METHERINGHAN R.A., "Measuring the Net Cumulative Coverage of a Print Compaign", Journal of Advertising Research, Vol. 4, December 1964, pp. 23-28.

(97) - HOFMANS P., "Measuring the Cumulative Net Coverage of Any Combination of Media", Journal of Marketing Research, vol. 3, August 1966, pp. 269-278.

(98) - BERNSTEIN A., "Computer Simulation of Media Exposure", A Report of the Sixth Meeting of the A.R.F. Operation Research Discussion Group, New-York, 1961.

(99) - Simulmatic Corporation, Simulmatic Media Mix, 1. General Description, 2. Technical Description, October 1962.

(100) - GENSH D.H., "A computer Simulation Model for Selecting Advertising Schedules". Journal of Marketing Research, Vol. 7, May 1969, pp. 203-214.

(101) - FRIEDMAN L., "Constructing a Media Simulation Model", Journal of Advertising Research, Vol. 10, August 1970, pp. 33-39.

(102) - INTERMARCO Limited, Advertising Planning by Media-Planex, 1971.

(103) - Dynamo-Media Scheduling Suit of Programs, University of London, Atlas Computing Service.

(104) - ALPERT L., "Measuring Effects Through Simulation", Proceeding of the 18 th Annual Conference of the A.R.F., November 1972, pp. 48-50.

(105) - YOUNG L.F., "Advertising Media Models : Their Uses and Limitations", in Computer Innovations in Marketing, American Management Association, 1970.

(106) - BERTIER P., DU JEU P., "Simulation des comportements aléatoires de lecture : Le Modèle SCAL", Metra, vol. 6,pp. 647-659.

(107) - BENSON P.H., "Bivariate Normal Distribution to Calculate Media Exposure", Journal of Advertising Research, Vol. 9, 1969, pp. 41-47.

(108) - BROADBENT S.R., "Regularity of Reading", Journal of Marketing Research, August 1964, pp. 50-58.

(109) - CHARNES A., COOPER W.W., LEARNER D.B., SNOW E., "Media Mixing for Distribution of Frequencies with Beta-like and lognormal Distributions of Frequencies", Management Science Report, n° 189, Carnegie Mellon.

(110) - EHRENBERG A.S.C., TWYMAN, "On measuring Television Audiences", Journal of the Royal Statistical Society. A - Vol. 130, 1967, pp. 1-59.

(111) - FRIEDMAN L., "Calculating TV Reach and Frequency", Journal of Advertising Research, Vol. II, August 1971.

(112) - GOODHART G.J., EHRENBERG A.S.C., "Duplication of Television Viewing Between and Within Channels", Journal of Marketing Research, Vol. 6 May 1969, pp. 169-178.

(113) - GREENE J., STOCK J.S., Advertising Reach and Frequency in Magazines : New Informations for Evaluating Magazine Advertising,New-York,

Reader Digest Association, 1967.

(114) - HEADER R.S., "Distribution Theory of Net Cumulative Audience Coverage and Exposure Frequency in Media Planning", Proceedings of 1971 Conference, American Marketing Association, pp. 230-234.

(115) - LIEBMAN L., "Modal Reach & Frequency : An Improved Estimation Method", 17 th Annual Conference of the A.R.F., November 1971.

(116) - KWEREL, S.M., "Estimating Unduplicated Audience and Exposure Distribution", Journal of Advertising Research, Vol. 9, June 1969, pp. 46-53.

(117) - COSMIC - A Comprehensive Press Planning System, Foote Cone and Belding, London, September 1970.

(118) - FRY C., "The Concept of Casualness", Feedback, n° 2, Apris 1973.

(119) - GENSCH D., Advertising Planning : Mathematical Models in Advertising Media Planning, Elsevier Scientific Publishing Co., Amsterdam, 1973.

CHAPITRE 3

LES OBJECTIFS DE L'ETUDE

Cette dissertation tend à la réalisation des six objectifs suivants :

- Le regroupement et la description de tous les modèles connus d'exposition aux supports publicitaires.
- La recherche et la classification de leurs hypothèses, explicites et implicites.
- La construction d'une grille d'évaluation à critères multiples permettant d'évaluer les mérites théoriques, empiriques et pratiques des modèles.
- La mise à l'épreuve des hypothèses et l'évaluation des performances des modèles à l'aide de banques de données unifiées.
- Le développement de nouveaux modèles d'exposition aux supports publicitaires.
- L'évaluation et le classement par ordre de mérite des différents modèles d'exposition aux supports publicitaires.

Les spécialistes de la planification publicitaire peuvent espérer de cette étude les bénéfices suivants : une meilleure appréhension du phénomène de l'exposition aux supports publicitaires, de meilleurs critères d'évaluation pour choisir parmi les modèles existants et de nouvelles directions de recherche pour l'établissement de meilleurs modèles.

1. Le Regroupement et la Description de tous les Modèles connus d'Exposition aux Supports publicitaires

Nous n'entendons pas nous limiter aux seuls modèles largement connus

et ayant fait l'objet de publication dans les revues spécialisées. Nous mettrons tout en oeuvre pour apporter des informations détaillées sur des modèles moins connus. Nous inclurons dans notre comparaison plusieurs modèles développés à l'étranger, en Angleterre, aux Etats-Unis et en Australie. Nous espérons ainsi remédier en partie au manque d'information qui règne dans la description des modèles et qui s'explique principalement par la valeur commerciale substantielle qui s'attache à un bon modèle d'exposition aux supports publicitaires. En conséquence, nos conjonctures et déductions viendront compléter, chaque fois que cela sera nécessaire, les lacunes présentes dans la description du modèle.

Chaque modèle sera décrit selon la même séquence : description des hypothèses, présentation des équations, estimation des paramètres, et sera suivi d'un exemple numérique décrivant la conduite des calculs. On examinera ensuite les performances empiriques du modèle sur la base de banques de données unifiées et on commentera les résultats.

2. La Recherche et la Classification des Hypothèses des modèles

On donnera un cadre de classification général basé sur la nature des hypothèses formulées au sujet des individus, des supports et des parutions. Les hypothèses de chaque modèle seront étudiées, et une attention spéciale sera accordée aux hypothèses implicites. Cet effort de clarification permettra de développer une meilleure compréhension des mécanismes de comportement sous-jacent à l'exposition aux supports publicitaires et révèlera des domaines inexplorés provenant de combinaison d'hypothèses jusqu'alors négligées.

3. La Construction d'une Grille d'Evaluation à Critères multiples

En évaluant chaque modèle nous consignerons les caractéristiques qu'il serait souhaitable qu'un modèle d'exposition satisfaisant possédât dans les domaines théorique, empirique et pratique. Les critères théoriques, dérivés des nécessités métathéoriques de cohérence, d'économie et de testabilité, seront complétés par les critères empiriques de robustesse, d'applicabilité et d'exactitude des prédictions, et par les critères pratiques liés à la simplicité et à la rapidité des calculs ainsi qu'à l'accessibilité des données requises. On construira ainsi trois grilles d'évaluation, une pour les modèles d'accumulation, une pour les modèles de duplication, et une autre pour les modèles d'exposition complets.

4. Le Test et l'Evaluation des Modèles à l'Aide de Banques de Données Unifiées

On utilisera trois banques de données : une pour tester les modèles d'accumulation, une pour tester les modèles de duplication et une pour tester les modèles d'exposition complets. La première banque de données provient de deux études réalisées par Politz (1), (2), qui sont considérées par les praticiens comme étant parmi les plus soigneusement effectuées à ce jour (3). Ces deux études mesurent les audiences cumulées jusqu'à six parutions pour un total de 10 magazines. Nous utiliserons ces données pour comparer la validité empirique de cinq modèles Ad Hoc d'accumulation et de onze modèles stochastiques d'accumulation. Cinq modèles de simulation seront testés sur la base des données recueillies en 1964 par Simmons (4) et sur la base de données simulées. Les meilleurs modèles d'accumulation seront ensuite testés contre les audiences cumulées jusqu'à 13 parutions successives mesurées par Politz pour les magazines Life (5) et Better Home and Gardens (6).

La seconde banque de données comprend les résultats de deux études de duplication faites, l'une aux Etats-Unis (7), et l'autre en France (8). L'étude Américaine couvre un total de onze magazines. Trois combinaisons de cinq magazines ainsi que six combinaisons de 11 magazines y sont tabulées. L'étude française couvre cinq magazines et toutes les combinaisons possibles ont été tabulées.

Au total, onze modèles de duplication d'audience sont testés et évalués sur la base de 26 plans publicitaires.

La troisième banque de données a été préparée spécialement pour notre recherche grâce à la coopération des sociétés Interactive Market Systems (9) et W.R. Simmons & Associates (10). Elle comporte 57 plans publicitaires test couvrant un vaste champ de combinaisons de magazines. Pour chaque plan test les audiences de chacun des magazines entrant dans la combinaison ont fait l'objet de tabulations croisées qui permettent de comparer les prédictions de 15 modèles d'exposition complets à la couverture nette réelle ainsi qu'à la distribution des occasions de voir. Cette banque de données constitue, à notre connaissance, la meilleure source d'information actuellement disponible pour tester les modèles d'exposition aux supports publicitaires. Pris dans leur ensemble, les 47 modèles d'exposition examinés dans notre recherche constituent la première étude comparative exhaustive réalisée dans le domaine de l'exposition aux supports publicitaires.

5. Le Développement de Nouveaux Modèles d'Exposition aux Supports Publicitaires

Un total de onze nouveaux modèles d'exposition est présenté dans cette dissertation. Cinq d'entre eux sont des modèles d'accumulation, quatre sont des modèles de duplication et quatre sont des modèles d'exposition complets. Dans les deux premières catégories de modèles, nous avons réussi à améliorer les résultats obtenus par les modèles préexistants. Dans la dernière catégorie, nous avons réussi à dépasser les performances des modèles publiés, mais nous n'avons pu surpasser les remarquables performances des deux meilleurs modèles couverts par le secret commercial.

6. L'Evaluation et le Classement des Modèles d'Exposition aux Supports Publicitaires.

Les grilles d'évaluation développées dans notre étude sont utilisées pour fournir une appréciation multicritère des 47 modèles d'exposition. Les modèles dominés sont indiqués et des conclusions sont présentées portant sur le bilan des recherches passées et sur l'évolution probable des recherches futures.

NOTES

(1) - POLITZ Alfred Research, Inc., A Study of Four Media, their Accumulative and Repeat Audiences, New-York, Time Inc., 1953

(2) - POLITZ Alfred Life Study of Consumer Expenditures, New York, Time Inc., 1957

(3) - SIMMONS W.R. "How to Measure Magazine Audiences", 19th Annual Conference of the A.R.F., New York, 1973

(4) - SIMMONS W.R. & Associates Research, Inc., 1964 Study of Selective Markets, New York, 1964

(5) - POLITZ Alfred Research, Inc., A Study of the Household Accumulative Audience of Life, New York, Time Inc., 1952

(6) - POLITZ Alfred Research, Inc., A 12 Months Study of Better Homes & Gardens Readers, Meredith Publishing Co., Des Moines, 1956

(7) - POLITZ Alfred Media studies, Main Report : 1964 Politz Magazine Study- Audiences of 11 Magazines, New York, February, 1965

(8) - MARC M Etude de la Couverture Réalisée par Plusieurs Revues Techniques sur une Clientèle Spéciale, IREP, Etude N° 6bis Paris, Novembre 1962

(9) - Interactive Market Systems, Inc., 19 West 44th Street, New York

(10) - SIMMONS W.R. & Associates Research, Inc., 235 East 42nd Street, New York

PREMIERE PARTIE

LES CONCEPTS DE BASE

CHAPITRE IV

DEFINITIONS ET MESURES

... aussi nous arrivons à la conclusion que les services de recherche syndiqués offrent à leurs clients une mesure inexacte d'une variable sans signification ...

William S. Blair

"The case against magazine audience measures"
Journal of Advertising Research, April 1974

Une grande confusion entoure l'utilisation de mots tels que : circulation, couverture, audience, exposition, couverture nette, fréquence, accumulation et duplication. Le sens de ces mots a eu tendance à varier à travers le temps, les pays, les supports publicitaires, et même d'une agence publicitaire à une autre.

Etant donné qu'une compréhension commune de la terminologie est nécessaire à la comparaison des différents modèles d'exposition présentés par la littérature, il est utile de s'en tenir à des définitions strictes. L'examen d'un certain nombre de compilations terminologiques par, Graham I. (1), Standard Rate & Data Service (2), Advertising Research Foundation, Audience Concept Committee (3), Brown I.O., Lessler R.S., Weilbacher W.M., (4), Sondage C.H., Fryburger V. (5), Lucas D.B., Britt S.H. (6), a révélé qu'aucune d'entre elles n'était totalement satisfaisante pour nos besoins, étant, soit quelque peu dépassée, soit incomplète ou inconsistante.

Pour chaque concept étudié, sa définition sera complétée par une description des procédures de mesure par lesquelles le concept devient opérationnel. Une revue des méthodes utilisées par les sociétés spécialisées dans la mesure des audiences sera présentée pour chaque concept. Chaque fois que différentes méthodes de mesure seront disponibles nous comparerons leurs points forts et leurs points faibles.

1. La Circulation

La circulation d'un support est le nombre d'unités physiques à travers lesquelles le message publicitaire est distribué (7).

a) Dans le cas de supports imprimés, la circulation est mesurée par le nombre moyen d'exemplaires mis en circulation à chaque parution. La moyenne est évaluée sur une période de six mois. Le chiffre moyen d'exemplaires distribués en six mois est fourni par chaque éditeur sous la forme d'un compte rendu assermenté bi-annuel.

11. Des données sur la circulation des supports imprimés existent depuis la création en 1914 aux USA de l'Audit Bureau of Circulation A.B.C. (8). Les données recueillies par A.B.C. comprennent le nombre total d'exemplaires imprimés, subdivisés en exemplaires vendus, en exemplaires promotionnels et en exemplaires non distribués. Les exemplaires vendus sont divisés en abonnements

et vente au détail (ex. : Kiosque à journaux). Les éditeurs de journaux et de magazines d'envergure nationale fournissent des comptes rendus géographiques par états et par comtés. Enfin, les abonnements sont analysés en comparant les proportions d'abonnés nouveaux et d'abonnés fidèles.

La qualité de ces données est excellente puisque les chiffres annoncés par les éditeurs sont vérifiés annuellement par les contrôleur A.B.C. Deux autres organisations, "Business Publication Audit of Circulation" et "Verified Audit Circulation Company" fournissent des données similaires pour les publications professionnelles.

La société "Standard Rate and Data Service" (9) publie mensuellement les circulations de 852 magazines d'intérêt général et de 200 revues agricoles dans un document intitulé "Consumer Magazine and Farm Publication Rates and Data".

12. En ce qui concerne les supports publicitaires radiodiffusés, la circulation est définie par le nombre de récepteurs radio ou télévision en fonctionnement pendant une période de temps déterminé. Il est important de noter que le cadre de référence qui est une période de 15 minutes pour les supports radiodiffusés ne se compare pas aisément avec la période de six mois retenue pour les supports imprimés. De plus,la circulation d'un support radiodiffusé varie considérablement selon l'heure de la journée, le jour de la semaine et la période de l'année. Les fluctuations de la circulation des supports imprimés sont en général beaucoup moins importantes.

La circulation atteinte par les principales chaînes de télévision est mesurée par la société A.C. Nielsen (10) et se présente sous la forme d'un index (Nielsen Television Index ou NTI en abrégé). L'index NTI mesure l'écoute quotidienne au moyen d'un compteur enregistreur (11) branché sur le récepteur de télévision de 1200 ménages répartis à travers les USA. Ces compteurs fournissent un enregistrement permanent, minute par minute, de toutes les chaînes regardées par les ménages, 48 semaines par an.

La société ARB (12) (American Research Bureau) collecte des données similaires au moyen de livres d'écoute tenus à jour par un échantillon de ménages. Le comité Harris (13) qui a enquêté sur la qualité de ces données a formulé deux critiques. Un grand nombre de foyers refusent de coopérer et doivent être remplacés. La représentativité de l'échantillon de ménages ne

tient pas compte du dernier recensement. Néanmoins, une étude financée par les trois chaînes de télévision et l'association nationale des stations de télévision a montré que les résultats obtenus par ARB et par Nielsen sont extrêmement proches.

Les données concernant les stations de radio ne sont pas disponibles de façon continue mais chaque station locale les fournit sur demande. La qualité de ces données est mal connue car on sait peu de chose sur les méthodes de mesures utilisées par chaque station ainsi que sur la taille des échantillons en cause. En conséquence, les comparaisons entre stations sont difficiles. La principale difficulté tient au fait que les récepteurs de radio en fonctionnement se trouvent aussi bien dans les voitures particulières que dans les foyers des personnes enquêtées.

13. En ce qui concerne la publicité extérieure, la circulation proprement dite est égale au nombre de supports physiques de l'affichage. Malheureusement, lorsque les spécialistes des média mentionnent la circulation de la publicité extérieure, ils se réfèrent au nombre total d'individus qui passent en vue des supports physiques d'affichage. Nous avons là un bon exemple de confusion entre le concept de circulation et le concept d'audience que nous allons définir plus loin. La circulation, au sens propre, de la publicité extérieure n'est pas mesurée de façon permanente. Toutefois, les diverses régies qui vendent des espaces publicitaires extérieurs sont à même de fournir des données concernant le nombre et la surface des divers points d'affichage pour les diverses villes qui entrent dans leur réseau.

En résumé, il convient de retenir que les données de circulation ne nous renseignent que sur le nombre de supports physiques des messages publicitaires, en état de fonctionnement pendant une période de temps déterminée. Elles ne nous renseignent jamais sur le nombre d'individus réellement en contact avec ces supports et sont par conséquent d'une utilité très limitée pour la planification d'une campagne publicitaire. En d'autres termes, la circulation n'est que la partie visible de l'iceberg, car chaque support physique est susceptible d'être vu, lu, écouté par plusieurs individus. La circulation ne mesure donc pas l'audience, elle en donne simplement la limite minimale. Il convient aussi de distinguer entre la circulation et la pénétration d'un support.

2. La Pénétration

La pénétration d'un support dans la population d'une zone géographique donnée est égale à la proportion de cette population possédant au moins une unité physique capable de transmettre les messages publicitaires véhiculés par ce support. Le concept de pénétration est utilisé essentiellement pour les supports radiodiffusés. Il diffère de la simple possession d'un appareil récepteur qui n'implique pas que le signal de la station étudiée, télévision ou radio, puisse être effectivement reçu. La possession d'un appareil récepteur ne représente que la limite supérieure de la pénétration. La pénétration réelle dépend essentiellement de facteurs géographiques et de la puissance en watts de la station émettrice ou réémettrice la plus proche. D'autre part, la différence entre la pénétration et la circulation est liée au taux d'utilisation de l'appareil récepteur. Ainsi, la pénétration d'un support radiodiffusé représente la limite supérieure de sa circulation.

On peut remarquer incidemment que, pour les supports imprimés, l'absence de toute mesure du taux d'utilisation des exemplaires ne nous permet pas de distinguer entre leur circulation et leur pénétration (14).

Les données de pénétration, plus encore que les données de circulation sont de peu d'utilité pour la planification publicitaire puisqu'elles ne nous renseignent que sur le nombre potentiel de supports physiques disponibles pour véhiculer le message publicitaire et non sur le nombre d'individus qui sont entrés en contact avec lui.

3. L'Exposition

Il convient de distinguer nettement entre deux types d'expositions; l'exposition au support publicitaire d'une part et l'exposition au message publicitaire d'autre part.

- Chaque fois qu'un individu rencontre un support, soit visuellement, soit auditivement, une exposition au support est censée s'être produite.
- Chaque fois qu'un individu rencontre un message publicitaire, soit visuellement, soit auditivement, une exposition au message est censée s'être produite.

Pour un type de support seulement, la publicité extérieure, l'exposition au support garantit presque sûrement l'exposition au message (15). Mais ceci n'est pas vrai pour tous les autres média. Le coeur du concept d'exposition

est l'existence d'une opportunité physique de voir ou d'entendre. Toutefois, l'existence de cette opportunité physique ne garantit pas qu'une perception visuelle ou auditive se soit en fait produite. En effet, la perception consciente requiert un minimum d'action volontaire d'encodage et de décodage communément appelée attention.

Notre recherche ne porte que sur le phénomène d'exposition aux supports. Nous ne traiterons pas de l'exposition aux messages ainsi que nous en sommes expliqués dans l'introduction.

Bien entendu, le nombre d'expositions générées par un support est en relation avec la circulation de ce support car plus nombreuses sont les unités physiques distribuées, plus nombreux sont les individus susceptibles d'entrer en contact avec elles. Mais il ne s'agit pas d'une relation bi-univoque. Le nombre d'expositions dépend du comportement des individus et de leur façon de "consommer" les divers supports. C'est précisément tout ce qui fait la difficulté de leur mesure car il n'est pas possible, en général, de recourir à l'observation directe.

Les techniques de mesure de l'exposition varient d'un média à un autre. Les différences reflètent à la fois la nature propre de chaque média et la concurrence qu'ils se livrent afin de s'attirer un maximum d'annonceurs. Ces techniques peuvent être classées en trois catégories :

- L'observation directe des comportements de lecture ou d'écoute.
- L'enregistrement, par la personne étudiée, de ses propres comportements de lecture ou d'écoute, à l'aide d'un journal d'écoute ou de lecture.
- Le recours au souvenir conservé par la personne de ses lectures ou écoutes récentes.

La première technique ne peut, malheureusement, dépasser le cadre des expériences de laboratoire, à la fois pour des raisons éthiques et pour des raisons de coût. En ce qui concerne la seconde technique, l'expérience montre que peu d'individus sont suffisamment motivés pour s'astreindre à enregistrer les expositions au moment même où elles se produisent. La plupart du temps le journal d'écoute ne sera rempli qu'en fin de journée, voire en fin de semaine, ce qui introduit le recours à la mémoire que l'on avait cru pouvoir éliminer. En conséquence, toutes les méthodes doivent recourir, à des degrès divers, à la mémoire des enquêtés. Ceci peut être réalisé matériellement par le recours à

des enquêtes personnelles à domicile, par des enquêtes téléphoniques ou par des questionnaires remplis par l'enquêté lui-même, avec ou sans la présence de l'enquêteur comme c'est le cas pour les enquêtes par voie postale.

31. Les Média imprimés

Nous examinerons tout d'abord les méthodes de mesure de l'audience des magazines, puis celle des journaux. Notre discussion portera sur les problèmes liés au choix de la technique de collecte, au choix du critère de lecture et au choix de la formulation des questions. Nous décrirons ensuite les diverses méthodes utilisées en France et aux USA, puis nous évaluerons l'ampleur des différences attribuables à la technique de collecte, au critère de lecture retenu et à la formulation des questions.

311. Le Choix de la technique de collecte des données

Sur le plan de la technique de collecte des données, on distinguera entre les méthodes ponctuelles et les méthodes longitudinales par panel. Dans la première catégorie, un individu est interrogé, à un moment donné, le plus souvent sous la forme d'interview personnel, sur ses lectures de la semaine ou du mois précédant le jour de l'interview. C'est la technique adoptée en France par le Centre d'étude des Supports de Publicité (CESP) (16) et aux Etats-Unis par Simmons (17).

Dans la seconde catégorie, un individu, contacté personnellement par un enquêteur et ayant accepté de participer, se voit remettre un journal de lecture dans lequel il devra consigner jour par jour les divers supports avec lesquels il est entré en contact. A la fin de chaque semaine le journal de lecture est retourné par la poste.

C'est la technique utilisée en France par Secodip (18) pour son panel media-produit.

Quelle que soit la technique choisie, on retrouve à des degrés divers les sources potentielles d'erreurs suivantes :

- Biais dû à la sélection des participants. En dehors de l'erreur d'échantillonnage purement aléatoire et qui peut être minimisée par l'utilisation d'un taux de sondage approprié, il existe un biais systématique dû au fait qu'un certain

nombre de participants désignés n'ont pu être contactés. En pratique, on palliera à l'absence des participants en prévoyant plusieurs retours à leur domicile. Toutefois, il est difficile d'apprécier le biais introduit par les refus systématiques de participation. Ce problème est nettement plus marqué en ce qui concerne le recrutement des panelistes et ne sera pas totalement résolu par les techniques de redressement d'échantillon qui consistent à donner des poids différents à chaque catégorie d'individus afin de reconstituer la représentativité de l'échantillon.

- Biais dû aux attentes conjointes du participant et de l'enquêteur.

La connaissance du but de l'enquête par l'enquêteur comme par l'enquêté conduit à de nombreuses distorsions. L'enquêté aura plutôt tendance, consciemment ou non, à gonfler ses comportements de lecture, spécialement en ce qui concerne les magazines connus ou prestigieux, simplement pour faire plaisir à l'enquêteur et pour s'éviter d'avoir à fouiller plus avant dans sa mémoire. Si la liste de magazines est particulièrement longue, ceux qui se trouvent en bout de liste seront sous-évalués car l'enquêteur comme l'enquêté seront fatigués et pressés d'en finir.

Souvent l'enquêté se réfèrera à son comportement de lecture habituel bien plus qu'à ses lectures réelles de la semaine précédente. Il énoncera ce qu'il "aurait dû lire" par rapport à ses habitudes plus que ce qu'il a réellement lu.

312. Problèmes liés au choix du critère de lecture

Il convient de préciser les critères utilisés pour distinguer un lecteur d'un non lecteur.

- Si l'on se contente d'une affirmation verbale de la part de l'enquêté, il s'agit d'un lecteur déclaré.
- Si, par contre, on demande à l'enquêté de décrire le numéro qu'il prétend avoir lu et qu'il y parvient, il s'agit d'un lecteur prouvé.
- Si l'on soutient la mémoire défaillante de l'enquêté en lui présentant plusieurs numéros et qu'il les reconnaît, il s'agit d'un lecteur reconnu.

Chacun de ces critères présente des avantages et des inconvénients.

Le critère du lecteur déclaré est le plus facile à appliquer. Il est même parfois le seul possible dans le cas des enquêtes par voie postale ou par panel. Son degré de distorsion est difficile à apprécier. On inclut parfois dans la liste des magazines présentés le nom d'un pseudo-magazine qui n'existe pas. Le pourcentage de personnes prétendant l'avoir lu donne une idée du degré de surestimantion susceptible d'affecter l'audience des titres les moins connus. Par contre cette technique ne nous renseigne pas sur le degré de surestimation propre aux titres connus et, spécialement, aux titres les plus prestigieux. Par ailleurs, il est fort probable que ce critère conduira à l'oubli des magazines feuilletés par hasard dans des circonstances peu marquantes.

Le critère du lecteur prouvé présente les mêmes inconvénients que le critère du lecteur déclaré puisque rien n'est fait pour pallier à d'éventuelles insuffisances de mémoire. De plus, il est parfaitement concevable qu'un individu se souvienne d'avoir lu un certain magazine mais soit totalement incapable de décrire le numéro ou son contenu. Enfin, il est possible que l'enquêteur, soucieux d'aider l'enquêté, lui fournisse lui-même les éléments de la réponse. De toute façon, la mise en oeuvre de ce critère sera délicate, la sous-estimation des titres les moins connus sera probablement accentuée, sans que l'on soit tout à fait certain d'avoir mis fin à la surestimation des titres connus.

Le critère du lecteur reconnu présente l'inconvénient d'obliger l'enquêteur à transporter un exemplaire de chacun des magazines à étudier. De plus, comme il n'est pas possible de faire parcourir à l'enquêté tous ces magazines, on doit se servir au préalable d'une question filtre dont le but est de déterminer une liste de magazines que l'enquêté pourrait avoir lus. Toutefois, l'existence même de cette question filtre conduit à un cercle vicieux car elle consiste à utiliser précisément le concept de lecture déclarée qu'il s'agissait d'éviter. Toutefois, mis à part les sous-estimations liées à la question filtre, la méthode permet de diminuer, dans une certaine mesure, la surestimation propre aux titres connus et prestigieux. Néanmoins, la durée de l'interview tend à devenir prohibitive, spécialement pour les enquêtés lecteurs de plusieurs magazines.

En pratique, on peut distinguer 3 niveaux de soutient mémoriel :

- Au premier niveau on présente à l'enquêté un paquet de cartes sur lesquelles figure le logotype de chacun des magazines étudiés.

C'est la solution adoptée en France par le CESP pour ses enquêtes ponctuelles et par Secodip pour ses panels media.

Aux Etats-Unis TGI (19) utilise la même procédure pour ses questionnaires d'audience, tandis que Simmons use la présentation des logotypes pour sa question filtre seulement.

- Au second niveau, on présente à l'enquêté uniquement la couverture du magazine. C'est la solution adoptée aux Etats-Unis par Starch (20).

- Au troisième niveau, on permet à l'enquêté de feuilleter le magazine qui a été préalablement allégé de la plupart des pages publicitaires. C'est la technique utilisée par Simmons, pour les individus ayant répondu par l'affirmative à la question filtre. Elle est connue sous le nom de "Through the Book" ou "Editorial Interest". Cette méthode a été mise au point dans les années 50 aux Etats-Unis par Alfred Politz Research (21). Sont définis lecteurs les enquêtés qui, après avoir feuilleté le magazine, déclarent être sûrs de l'avoir déjà eu entre leurs mains.

Le degré de sous-estimation de l'audience due à la présence de la question filtre a été récemment examiné par T. Joyce (22). Un échantillon de 750 individus a été soumis à la question filtre par la présentation des logotypes de 71 magazines. Ensuite, tous les individus, qu'ils aient déclaré ou non être lecteur, ont feuilleté en présence de l'enquêteur 4 magazines, en vue de déterminer s'ils reconnaissaient l'avoir lu ou non.

Seulement 457 personnes répondirent positivement à la question filtre, tandis que 575 reconnurent les numéros feuilletés. Il semble donc que la question filtre ait conduit à une sous-estimation de l'audience d'environ 20%.

Toutefois, le type de question posé ainsi que le critère retenu pour la définition d'un lecteur ne suffisent pas à définir entièrement une méthode de mesure d'audience. Il convient d'examiner maintenant la formulation de la question.

313. Problèmes liés au choix de la formulation des questions

Il convient de distinguer entre les questions se rapportant à la lecture d'un numéro donné du magazine, les questions se rapportant à la lecture du magazine pendant une période donnée et enfin les questions d'habitude de lecture. Supposons un mensuel pour lequel la durée de vie moyenne d'un numéro serait de trois mois. Cela signifie qu'un numéro de plus de 3 mois d'âge n'est plus lu par personne. Supposons en plus que la moitié des expositions à ce

mensuel aient lieu durant le premier mois, le tiers durant le deuxième mois et le restant durant le troisième mois. Dans ces conditions, l'évolution du nombre de lecteurs par mois et par numéro peut être représentée par le diagramme suivant :

Diagramme de Lecture

Numéro du mois de	Audience	Janvier	Février	Mars	Avril	Mai
JANVIER	A	50 % A	33 % A	17 % A	-	-
FEVRIER	B	-	50 % B	33 % B	17 % B	-
MARS	C	-	-	50 % C	33 % C	17 % C
AVRIL	D	-	-	-	50 % D	33 % D
MAI	E	-	-	-	-	50 % E

Si l'on admet en première analyse que l'audience des numéros reste constante d'un mois à l'autre, c'est-à-dire que A=B=C=D=E, le diagramme précédent montre qu'une question portant sur la lecture d'un numéro donné, disons Janvier, donnera la même audience qu'une question portant sur les lectures d'un mois donné, disons Avril.

Toutefois, les audiences sont soumises à la fois à des variations saisonnières régulières et à des fluctuations accidentelles liées à la survenance d'évènements nationaux ou internationaux de grande importance. Du fait qu'elle recouvre plusieurs numéros, la question portant sur la lecture d'un mois donné sera moins sensible aux fluctuations accidentelles que la question portant sur un numéro donné. Par contre, elle sera plus sensible à la variation saisonnière caractérisant le mois choisi.

La question se rapportant à la lecture du numéro de Janvier ne pourra être posée qu'au début du mois d'Avril afin de prendre en compte les 17 % de lecteurs tardifs. A cette date, 50 % des personnes seront interrogées au sujet d'une lecture ayant eu lieu il y a trois mois et dont le souvenir sera nécessairement très estompé. Le risque de sous-estimation de l'audience est donc très net. La question relative aux lectures du mois précédent réduit la distorsion due à l'oubli mais par contre en introduit deux nouvelles dues aux lectures parallèles de ceux qui lisent en Avril à la fois le numéro de Mars et celui

d'Avril et aux lectures répétées de ceux qui lisent deux mois de suite le même numéro.

Afin d'illustrer ces deux phénomènes, la table suivante décrit les comportements de lecture d'un hebdomadaire par 3 lecteurs fictifs.

LECTURES REGULIERES, PARALLELES ET REPETEES

A - Lectures Régulières

Semaine		1	2	3	4
Numéros	1	1	0	0	0
	2		0	0	0
	3			1	0
	4				0

Lectures de la dernière semaine = 1/2

Numéros lus = 2/4

B - Lectures Parallèles

Semaine		1	2	3	4
Numéros	1	0	1	0	0
	2		1	0	0
	3			0	1
	4				1

Lecture de la dernière semaine = 1/2

Numéros lus = 4/4

C - Lectures Répétées

Semaine		1	2	3	4
Numéros	1	1	1	0	0
	2		0	0	0
	3			1	1
	4				0

Lecture de la dernière semaine = 4/4

Numéros lus = 2/4

- Le lecteur A lit un numéro toutes les deux semaines, ne lit jamais plus d'un numéro chaque semaine, et ne lit jamais un même numéro deux semaines de suite. En conséquence, pour une semaine quelconque, il y a une chance sur deux qu'il soit compté comme lecteur, ce qui concorde parfaitement avec le fait que cette personne lit deux numéros sur 4.

- Le lecteur B lit toutes les deux semaines, les semaines où il lit, il lit 2 numéros différents, et il ne lit jamais un même numéro deux semaines de suite. Il ne sera compté comme lecteur que les semaines 2 et 4, donc en moyenne une semaine sur deux, alors qu'en fait il lit bien 4 numéros sur 4. Donc, en cas de lectures parallèles, la méthode de la lecture de la dernière semaine sous-estime l'audience véritable.

- Le lecteur C lit le même numéro deux semaines de suite, ne lit jamais plus d'un numéro par semaine, et lit un numéro sur deux. Quelle que soit la semaine où on l'interroge, il sera donc compté comme lecteur, alors qu'en fait il ne lit que 2 numéros sur 4. Donc, en cas de lectures répétées, la lecture de la dernière semaine surestime l'audience véritable.

Théoriquement, le biais introduit par les lectures parallèles et répétées pourrait être éliminé en demandant à l'enquêté de distinguer entre la première lecture et les lectures répétées d'un numéro donné. Toutefois, l'expérience tend à prouver que cette information est bien au-delà de ce que l'on peut attendre des facultés de mémorisation d'un enquêté moyen.

De plus, on ne sait pas dans quelle mesure la sous-estimation due aux lectures parallèles est compensée par la surestimation due aux lectures répétées. Il est probable, néanmoins, que l'influence des lectures répétées soit plus grande pour les mensuels que pour les hebdomadaires. Les difficultés propres à la mesure des lecteurs d'un numéro donné ou d'une période donnée ont suscité un troisième type de question.

Au lieu d'interroger l'individu sur ses lectures d'un numéro donné ou sur ses lectures d'une période donnée, on peut l'interroger sur ses habitudes de lecture en lui proposant de cocher, sur échelle de fréquence verbale ou numérique, la position qui correspond le mieux à ses habitudes de lecture du magazine. En France, l'échelle d'habitude utilisée par le CESP jusqu'en 1970 était du type numérique :

CESP (jusqu'en 1970) :

" Voici un paquet de cartes représentant chacune une revue paraissant tous les mois (semaines pour les hebdomadaires). Pour chacune d'elles, voulez-vous me dire si, personnellement, vous la lisez ou simplement la feuilletez CHEZ VOUS ou DEHORS :

Mensuels	Hebdomadaires
- Tous les mois	- Toutes les semaines
- 6 à 10 fois par an	- 2 à 3 fois par semaine
- 3 à 4 fois par an	- 1 fois par mois
- 1 à 2 fois par an	- tous les 2 à 3 mois
- ni lu ni feuilleté depuis un an	- 1 à 2 fois par an
	- ni lu ni feuilleté depuis un an".

Depuis 1971 l'échelle d'habitude de CESP est devenue verbale :

CESP (depuis 1971)

" Voici un paquet de cartes ... idem ... AU DEHORS :

Mensuels et Hebdomadaires

- Régulièrement
- Assez souvent
- Rarement
- Presque jamais
- Absolument jamais"

La question concernant la lecture de la dernière prériode est posée après la question d'habitude de lecture et pour tous les titres, sauf ceux pour lesquels l'enquêté a répondu "Absolument jamais". Elle a fait, elle aussi, l'objet de deux formulations différentes :

CESP (jusqu'en 1972) :

"Au cours des 30 (7 pour les hebdomadaires) derniers jours jusqu'à hier compris, avez-vous personnellement, lu ou feuilleté un numéro de ..., chez vous ou au dehors, même si c'était un numéro ancien ?"

oui ☐ non ☐

CESP (depuis 1973) :

"Quand, pour la dernière fois, avez-vous personnellement lu ou feuilleté un numéro de ..., chez vous ou au dehors, même un numéro ancien ?

Etait-ce
- Hier
- Il y a moins d'une semaine
- Il y a moins de 15 jours
- Il y a moins d'un mois
- Il y a moins de deux mois
- Il y a moins de trois mois
- Il y a plus de trois mois ?"

L'effet des quatre combinaisons possibles des deux types de questions, sur les habitudes et les lectures de la dernière période, a été étudié par Daniel Mugnier (23). En ce qui concerne les habitudes de lecture, il apparaît que la formulation "tous les mois" entraîne un pourcentage de réponses légèrement supérieur à celui obtenu par la formulation verbale "régulièrement". Par contre, les lecteurs occasionnels de la formulation verbale "rarement" et "presque jamais" sont beaucoup plus nombreux que ceux de la formulation numérique. De plus, la formulation verbale des habitudes semble être plus sensible aux variations saisonnières. Enfin, les différences entre les deux formulations sont plus marquées pour les mensuels que pour les hebdomadaires.

En conclusion, la formulation verbale tend à enfler les lecteurs occasionnels par rapport à la formulation numérique, et ce d'une façon plus marquée pour les mensuels, ainsi que le montre le tableau suivant :

Effets de la Formulation sur les Habitudes de lecture Déclarées

(% cumulé de lecteurs, moyenne de deux échantillons de 900 personnes)

Formulation Numérique		Formulation Verbale	
Hebdomadaires			
Toutes les semaines	4,75	Régulièrement	4,15
+ 2 à 3 fois par mois	6,50	+ Assez souvent	8,10
+ 1 fois par mois	9,50	+ Rarement	13,75
+ Tous les 2, 3 mois	12,45	+ Presque jamais	17,05
Mensuels			
Tous les mois	4,00	Régulièrement	2,55
+ 6 à 10 fois par an	5,95	+ Assez souvent	5,55
+ 3 à 4 fois par an	9,20	+ Rarement	11,75
+ 1 à 2 fois par an	13,05	+ Presque jamais	14,65

Source : Daniel Mugnier (23)

En ce qui concerne les lectures de la dernière période, la formulation verbale des habitudes de lecture tend à les augmenter par rapport à la formulation numérique et ce aussi bien pour les hebdomadaires que pour les mensuels. Ce phénomène est plus marqué pour l'ancienne formulation des lectures de dernière période que pour la nouvelle formulation ainsi que le montre le tableau suivant :

Effet combiné des formulations des questions d'habitude et de lecture de dernière période

Hebdomadaires		Habitudes de lecture		
		Formulation numérique	Formulation verbale	Moyenne
Lecture de la dernière période	Ancienne question	7,13	7,35	7,24
	Nouvelle question	5,52	5,72	5,62
	Moyenne	6,32	6,53	
Mensuels		**Habitudes de lecture**		
		Formulation numérique	Formulation verbale	Moyenne
Lecture de la dernière période	Ancienne question	6,82	7,74	7,28
	Nouvelle question	6,46	6,81	6,63
	Moyenne	6,64	7,27	

Source : Daniel Mugnier (23)

Si on met à part les distorsions liées à la formulation de la question d'habitude et si l'on admet en première analyse l'absence de distorsions mémorielles, il est possible de calculer l'audience d'un numéro moyen à partir de la répartition des individus entre les diverses classes de fréquence, ainsi que le montre l'exemple fictif suivant.

Echelle de Fréquentation	Individus	Probabilités	Audience Moyenne
Toutes les semaines	6	1	6
3 fois par mois	4	3/4	+3
2 fois par mois	4	1/2	+2
1 fois par mois	4	1/4	+1
ni lu, ni feuilleté	82	0	+0
Total	100		12

Toutefois, les études empiriques de Corlett et Osborne (24), Durand (25), Green (26), Schreiber (27), ont montré qu'une transformation directe des fréquences de lecture en probabilités conduit à une surestimation de l'audience. Ceci est dû à deux raisons.

Tout d'abord, lorsque l'on remplace les véritables (et inconnues) probabilités de lectures des individus, qui varient de façon continue entre zéro et un, par un nombre limité de probabilités discrètes (5 ans dans notre exemple), il se produit un "effet de découpage" qui diminue l'audience de façon systématique. Si nous notons m_k le moment non centré d'ordre k de la vraie distribution des probabilités individuelles de lecture, la valeur du biais systématique dans l'audience cumulée de N numéros sera égale à :

$$\text{Biais pour N numéros} = \sum_{k=0}^{k=N} (-1)^k \binom{N}{K} (m_k - m_1^k)$$

Il est difficile de remédier à l'effet de découpage car on ne peut guère augmenter le nombre de classes de fréquence au-delà de 5 ou 6.

La seconde cause de surestimation est due aux lectures répétées. Supposons qu'un individu ait lu deux fois le même numéro d'un hebdomadaire durant le dernier mois et rien d'autre.

Sa fréquence de lecture est-elle 1 fois par mois ou 2 fois par mois ? L'énoncé de la question CESP ne permet pas de trancher. Préciser dans l'énoncé qu'il s'agit d'indiquer la fréquence de lecture de numéros différents ne résoudrait rien car il est peu probable que l'enquêté soit capable de distinguer, avec précision entre deux lectures d'un même numéro et la lecture de deux numéros différents.

Pour remédier à cette surestimation, on a proposé d'étalonner la question d'habitude de lecture à l'aide des résultats de la question sur les lectures de la dernière période. En effet, il est facile de calculer, pour chacune des classes d'habitude séparément, le pourcentage d'individus qui déclarent avoir lu dans la dernière période. Ce pourcentage est alors pris comme probabilité de lecture associé à la classe d'habitude. Ce faisant, l'audience d'un numéro moyen, calculée par l'application de ces probabilités, sera nécessairement égale à l'audience mesurée par les lectures de la dernière période. Cette technique d'étalonnage est utilisée par T.G.I. pour le calcul des audiences cumulées. Toutefois, rien ne prouve que les lectures déclarées de la dernière période soient la vraie mesure de l'audience.

En fait, lorsque l'on compare les lecteurs reconnus d'un numéro avec les lecteurs déclarés de la dernière période, on constate que les premiers sont toujours moins nombreux que les seconds et que l'écart est plus marqué pour les mensuels que pour les hebdomadaires.

Ce phénomène a fait l'objet de nombreuses vérifications empiriques. Le tableau suivant présente les résultats obtenus par Politz (28) en 1967, Simmons (29) en 1969, et plus récemment, par Mugnier (23) en 1972.

Comparaison entre Lecteurs Reconnus et Lecteurs Déclarés

	Lecteurs Reconnus	Lecteurs Déclarés	Différence %
Politz 1967	(Lecture d'un numéro)	(Dernière période)	
Hebdomadaires et Bimensuels	168	181	+ 7,7 %
Mensuels	203	234	+15,3 %
Moyenne pour 14 Magazines	188	211	+12,2 %
Simmons 1969			
5 Magazines Féminins	208	242	+16,6 %
5 Magazines Généraux	414	462	+11,5 %
Reader's Digest	529	654	+25,5 %
Mugnier 1972	(Moyenne sur 2 numéros)		
3 Hebdomadaires			+17,6 %
3 Mensuels			+42,8 %

314. Les Méthodes utilisées en France et aux Etats-Unis

Si l'on combinait les deux techniques de collecte : ponctuelle ou panel, avec les trois formulations de la question : numéro, période ou habitude, et les trois critères de lecture : déclarée, prouvée, reconnue, on obtiendrait un total de 18 ($2*3*3$) méthodes de mesure des audiences des magazines.

Toutefois, on s'accorde à privilégier la formulation de la question. On distinguera donc trois familles de méthodes : la lecture au numéro, la lecture de dernière période et les habitudes de lecture, chacune d'elles présentant plusieurs variantes selon la technique de collecte des données et le critère de lecture retenu.

Le tableau suivant s'efforce de donner une vue d'ensemble des principales caractéristiques de chacune des méthodes et de la manière dont elles sont utilisées en France et aux Etats-Unis.

La MESURE de l'AUDIENCE des MAGAZINES en FRANCE et aux ETATS-UNIS

	Lecture au Numéro (through the book)	Lecture Dernière Période (Recent Reading)	Habitudes de Lecture (Reading Habit)
Société d'étude et variantes utilisées	Politz (Numéro entier) Simmons (Numéro squelette) Starch (Couverture seulement)	T.G.I. (logotypes) C.E.S.P. (logotypes) SECODIP (logotypes)	T.G.I. (Echelle numérique) C.E.S.P. (Echelle verbale depuis 1972)
Taille de l'échantillon	Politz = 8 000 Simmons = 20 000 Starch = 30 500	T.G.I. = 30 000 C.E.S.P. = 12 000 SECODIP = 2 panels de 4 500	
Méthode de sondage	Politz, Starch = stratifié et proportionnel Simmons = stratifié et non proportionnel	T.G.I. = stratifié et non proportionnel C.E.S.P. = aléatoire à 2 degrés SECODIP = échantillon constant	
Taux de réponse ou de participation	Simmons = 60 %	T.G.I. = 50 % C.E.S.P. = 70 % SECODIP = 50 %	
Nombre de vagues	Simmons = 2	T.G.I = 4 C.E.S.P. = 4 SECODIP = panel continu Avril-Juin	
Mesures répétées	Simmons = Chaque individu interrogé 2 fois	T.G.I. et C.E.S.P. = chaque individu n'est interrogé qu'une fois SECODIP = mesures hebdomadaires	
Nombre de magazines étudiés	Politz = 11 Simmons = 67 Starch = 58	T.G.I. = 140 C.E.S.P. = 67 SECODIP = 41	
Mise en oeuvre (questionnaire rempli par)	Politz, Simmons Starch = l'enquêteur	T.G.I. = l'enquêté C.E.S.P. = l'enquêteur SECODIP = le paneliste	
Critère de lecture	Politz, Simmons, Starch = Lecteur reconnu	T.G.I., C.E.S.P., SECODIP = lecteur déclaré.	

Maintenant que nous avons recensé les diverses méthodes de mesure de l'audience des magazines et discuté leurs avantages et inconvénients réciproques, il est intéressant de comparer leurs résultats.

315. Différences attribuables à la technique de collecte des données

a) Le simple fait de conduire une enquête à domicile plutôt que par voie postale introduit une distorsion appréciable lorsqu'il s'agit d'estimer l'audience d'un magazine mensuel.

Ainsi, lorsqu'en 1969 Stafco conduisit une enquête à domicile parmi 951 panelistes, l'audience de 20 mensuels mesurée par l'enquête à domicile fut en moyenne supérieure de 32% à l'audience mesurée par voie postale sur ces mêmes panelistes.

b) Lorsque la formulation de la question ainsi que la période couverte sont rigoureusement identiques, la différence entre les résultats du panel et ceux de l'enquête ponctuelle est minimum pour les magazines hebdomadaires ainsi que le montre la comparaison suivante :

Audience moyenne de		Panel Attwood	Enquête Ponctuelle IPA
5 hebdomadaires	1964	21 %	21,3 %
anglais	1965	19,1 %	19,4 %

Source : Parfitt (30)

La même conclusion peut être tirée de la comparaison du panel Secodip et de l'enquête C.E.S.P. pour l'année 1969,

Audience totale de 18 magazines hebdomadaires français	Panel Secodip	Enquête C.E.S.P.
	152,6	156,3

Source : Lancestre et Pinet (31)

c) Les habitudes de lecture déclarées sont très sensibles au mode de collecte. Les lecteurs réguliers apparaissent plus nombreux en cas d'enquête qu'en cas de panel, par contre les lecteurs irréguliers et les non-lecteurs apparaissent moins nombreux.

Déclarent lire	Enquête	Panel
4 numéros sur 4 (Réguliers	147	71
1 à 3 numéros sur 4 (Irréguliers)	141	157
non lecteurs	312	372

Source : Joyce et Bird (32)

Une conclusion similaire peut être tirée pour la France de la comparaison entre les habitudes de lectures mesurées par le panel Secodip, d'une part, et l'enquête CESP d'autre part. L'écart entre le panel et l'enquête s'accroît lorsque l'on descend l'échelle des fréquences de lecture.

Habitudes de Lecture	Panel SECODIP	Enquête CESP	Panel/Enquête
20 Hebdomadaires - 1967			
Régulièrement	105	125	83 %
Fréquemment	67	97	69 %
Infréquemment	80	177	45 %
12 Mensuels - 1967			
Régulièrement	54	58	93 %
Fréquemment	34	51	67 %
Infréquemment	40	111	36 %

Source : Pinet et Déribéré-Desgardes (33)

Il convient toutefois de noter que l'audience des mensuels est mesurée par Secodip à l'aide d'une question sur les lectures de la dernière semaine suivie d'une accumulation sur 4 semaines, tandis que la question CESP porte directement sur les lectures du dernier mois.

316. Différences Attribuables à la Formulation des Questions et aux Critères de Lecture

Les variations d'audience relatives au critère de lecture retenu ayant déjà été passées en revue, nous étudierons ici l'effet combiné de la formulation et du critère de lecture. Nous comparerons donc les audiences mesurées par la lecture au numéro sur la base des lecteurs reconnus, avec les audiences mesurées par la lecture de la dernière période sur la base des lecteurs déclarés. Comme la lecture au numéro n'est pas utilisée en France, les résultats présentés proviennent tous des Etats-Unis.

	Lecture au Numéro et Lecteurs Reconnus		Lecture de la Dernière Période et Lecteurs Déclarés
Année 1965	Polits (N. entiers)	Simmons (N. squelettes)	SR&DS
Better Homes & Gardens	19,200	17,408	21,451
Family Circle	15,550	12,435	16,536
Good Housekeeping	15,080	14,200	17,013
Ladies' Home Journal	14,770	13,871	15,972
Life	29,550	31,859	33,693
Look	28,110	29,303	27,800
MacCall's	21,750	16,598	20,221
Reader's Digest	35,770	37,395	35,930
Redbook	9,600	8,232	11,148
Saturday Evening Post	23,370	22,893	19,795
Woman's Day	13,310	10,957	15,543
	Lecture au Numéro Lecteurs Reconnus	Lecture Dernière Période & Lecteurs Déclarés	Pourcentage de Différence
Année 1972	Simmons	T.G.I.	
Better Homes & Gardens	22,724	31,630	39,1
Family Circle	14,516	23,890	39,2
Family Weekly	16,783	24,600	46,6
Good Housekeeping	15,845	24,340	53,6
Ladies' Home Journal	16,634	19,380	16,5
MacCall's	17,653	21,100	19,5
National Geographic	15,842	20,760	30,5
	Lecture au Numéro Lecteurs Reconnus	Lecture Dernière Période & Lecteurs Déclarés	Pourcentage de Différence
Année 1972	Simmons	T.G.I.	
Newsweek	13,690	17,150	25,3
Parade	32,417	29,030	-11,7
Playboy	16,377	20,880	27,5
Reader's Digest	42,616	40,930	-4,1
Redbook	10,473	13,330	27,3
Time	20,355	16,290	24,9
TV Guide	35,821	49,920	39,3
Woman's Day	14,174	16,810	18,6

Il ressort de cette comparaison que les audiences mesurées sur la base des lecteurs reconnus d'un numéro sont, en moyenne, inférieures aux audiences mesurées sur la base des lecteurs déclarés de la dernière période, ainsi que le montre la table suivante.

Comparaison entre les Lecteurs Reconnus et Lecteurs Déclarés de la Dernière Période

	Rapport Moyen	Rapport Minimum	Rapport Maximum
SR&DS Politz (1965)	104	85	117
SR&DS Simmons (1965)	109	95	142
T.G.I. Simmons (1972)	121	90	154

Les écarts constatés en 1965 se sont accrus en 1972. Parmi les 62 magazines étudiés en commun par T.G.I. et Simmons en 1972, 43 d'entre eux présentaient des différences supérieures ou égales à un tiers de l'audience.

Par contre, les différences entre les deux études utilisant la méthode des lecteurs reconnus d'un numéro est faible. Le rapport moyen Simmons/Politz s'établit à 95 avec un minimum de 76 et un maximum de 108. L'utilisation par Simmons de numéros amputés de leurs pages publicitaires paraît donc conduire à une sous-estimation de l'audience de 5 % en moyenne.

En définitive, les divers écarts constatés ne permettent pas de conclure à la supériorité définitive de l'une des méthodes en présence. On retiendra essentiellement que la lecture déclarée de la dernière période est plus sensible aux lecteurs secondaires et occasionnels, tandis que la lecture reconnue au numéro est plus sensible aux lecteurs primaires et réguliers. Par delà les problèmes soulevés par le choix d'un critère de lecture et d'une formulation de la question, il convient de mentionner deux problèmes supplémentaires. Le premier réside dans la détérioration continue du taux de réponse achevé auprès des enquêtés. Cette détérioration nuit évidemment à la représentativité de l'échantillon. Le second problème est celui de la mise à jour des audiences dans l'intervalle de temps séparant deux enquêtes. En effet, au premier délai séparant le moment de l'enquête de la publication des résultats s'ajoute un second délai créé par la périodicité annuelle des enquêtes. De plus, les achats d'espaces doivent être planifiés bien avant le début réel de la campagne si bien qu'il n'est pas rare qu'une campagne soit préparée sur la base

de données d'audience qui ont plus d'un an d'âge lorsque la campagne débute et plus de deux ans lorsqu'elle se termine.

Nous tournons maintenant notre attention vers la mesure de l'audience des journaux. Ce media qui, tant en France qu'aux Etats-Unis, reçoit la plus grande part des budgets publicitaires, est comparativement moins connu. Ce manque d'information, spécialement sensible aux Etats-Unis, tient d'une part au grand nombre des journaux en circulation et à leur caractère régional et d'autre part au sentiment que les mesures d'audience sont superflues puisque quasiment tout le monde lit un journal quotidien. Dans les années 50, on ne disposait guère que des diverses études, financées pour leur propre compte, par les plus grands des quotidiens. Depuis 1961, le Newspaper Advertising Bureau (34) mesure de façon régulière le niveau général de fréquentation des journaux quotidiens considérés dans leur ensemble en tant que medium unique et non pas titre par titre. En France, par contre, le Centre d'Etude des Supports de Publicité mesure l'audience de huit grands quotidiens parisiens de diffusion nationale (35). Aux USA, les bureaux d'études d'audience ne s'intéressent aux journaux quotidiens que depuis 1970 pour Simmons et 1973 pour TGI, encore ne mesurent-ils que l'audience du médium et non celles de quotidiens particuliers. La méthode de mesure utilisée est la lecture de la dernière période, modifiée à l'aide d'une séquence de questions indirectes visant à minimiser le biais en faveur des quotidiens les plus prestigieux. La formulation des questions utilisée par Simmons est la suivante .

On commence par demander à l'enquêté : "quels sont les journaux quotidiens que vous avez lu ou parcouru au cours de la semaine dernière, chez vous ou en dehors de votre domicile ?" La référence à une période d'une semaine, supérieure à la période de qualification qui est la lecture du jour précédent, fournit à l'enquêté l'occasion de mentionner le nom d'un des quotidiens les plus prestigieux, même si en fait il ne le lit pas. On lui demande ensuite : "mis à part aujourd'hui, quand avez-vous lu ou parcouru pour la dernière fois le journal X ?". Seules les personnes qui répondent avoir lu hier sont comptées comme lecteurs du journal en question. La même procédure est appliquée séparément par les éditions du Dimanche. Puisque l'audience des quotidiens tout comme celle des magazines s'accumule au cours du temps, chaque personne est enquêtée deux fois. La seconde enquête est utilisée pour calculer l'audience moyenne d'une parution et le taux d'accumulation. Les données sont divisées selon des critères socio-démographiques, géographiques et sociologiques. Les informations concernant la duplication entre l'audience des journaux est obtenue en

comptabilisant séparément les personnes qui ne lisent qu'un quotidien et qu'un seul, et celles qui en lisent au moins deux.

TGI affine cette analyse en mesurant aussi l'audience propre des diverses sections d'un quotidien type, telles que les petites annonces, les bandes dessinées, le sport, la mode, etc,... La table suivante permet de comparer les estimations du Newspaper Advertising Bureau avec celles de Simmons et de TGI.

Audience Adulte des Journaux Quotidiens
(USA-1972 - Milliers de Lecteurs)

Audience Moyenne d'un jour de semaine	Newspaper Advertising Bureau	Simmons	T.G.I.
Lit au moins un quotidien	101 356	98 003	93 393
Lit exactement un quotidien	72 976	71 554	
Lit deux quotidiens ou plus	28 380	26 449	
Audience cumulée d'une semaine	117 283	122 251	
Audience Moyenne du Week end			
Lit au moins un quotidien	82 927	90 088	91 313
Lit exactement un quotidien		72 827	
Lit deux quotidiens ou plus		15 261	

Le fait que les estimations soient très proches les unes des autres peut être attribué à l'emploi d'une méthodologie identique couplée avec une très large base de sondage obtenue en mesurant le medium en lui-même plutôt que des quotidiens particuliers.

32. Les Média Radiodiffusés

Pour les média radiodiffusés, le problème méthodologique réside dans la transformation des données de circulation, qui peuvent être enregistrées automatiquement mais qui n'indiquent que le nombre de récepteurs en fonctionnement à un instant donné, en données d'audience. Ceci est particulièrement difficile du fait que l'exposition aux supports radiodiffusés est le résultat d'un comportement passif, à l'opposé de la lecture qui nécessite une

participation active de l'individu. En conséquence, le souvenir de cette exposition sera probablement plus fugitif. De plus, il sera le plus souvent impossible de soutenir une mémoire défaillante par une réexposition au support comme on peut le faire avec un magazine ou avec un quotidien. Dans ces conditions, le recours à l'observation directe et aux journaux d'écoute s'avère indispensable.

Parmi les divers systèmes qui ont été suggérés en vue d'obtenir un dénombrement direct des personnes regardant le téléviseur familial, le plus connu est le Dynascope (36). Il comprend une caméra automatique photographiant à travers un miroir, à la fois les individus présents dans la pièce et l'écran du téléviseur. Les photographies sont prises à intervalle régulier dès que le récepteur est en fonctionnement. Elles permettent de déterminer le nombre de personnes présentes dans la pièce, ainsi que le nombre de personnes attentives. Toutefois, le Dynascope n'a jamais quitté le stade expérimental, car d'une part le codage des informations à partir des photographies est extrêmement fastidieux et d'autre part les messages publicitaires qui apparaissent dans l'intervalle de deux photographies ne sont pas pris en compte. De plus, les individus placés dans les coins de la pièce ou trop près du téléviseur sortent du champ photographique. Finalement, la présence du système est susceptible de modifier les habitudes d'écoute, pour ne rien dire de l'atteinte à la vie privée que représente un tel système de mesure.

Un autre effort d'observation directe a été utilisé par Television Advertising Representatives Inc. Quelques trois cents adolescents ont été entrainés à observer et à enregistrer les habitudes d'écoute télévisée des membres adultes de leur famille. Toutefois, les bureaux d'études spécialisés dans la mesure des audiences télévisées s'en tiennent à la méthode des journaux d'écoute.

La société Nielsen (10) a perfectionné cette méthode au cours d'un grand nombre d'années. Le journal hebdomadaire d'écoute est complété par un système d'enregistrement automatique, le Recordimeter, branché sur le récepteur de télévision. Le Recordimeter enregistre le nombre de demi-heures de fonctionnement du récepteur ainsi que la ou les chaines écoutées. Les versions initiales comportaient de plus une sonnerie et un signal lumineux chargé de rappeler à la ménagère, la nécessité de mettre à jour son journal d'écoute! Le double enregistrement permettait d'écarter les journaux d'écoute dont les informations ne correspondaient pas à celles du recordimeter. A partir de ces journaux,

Nielsen compile les cinq index d'audience suivants (37) :

1. Couverture : Le nombre de foyers susceptibles de recevoir une chaîne donnée. La couverture mesure la pénétration de chaque chaîne.

2. Foyers regardant la télévision : Le pourcentage de foyers dans lesquels, au moins, un téléviseur est en fonctionnement pendant une minute moyenne à une période déterminée de la journée. Ce pourcentage est calculé pour 230 locations géographiques différentes. Le deuxième index mesure la circulation du médium.

3. L'Audience totale : Le nombre de foyers écoutant une chaîne donnée, à une période déterminée de la journée, pendant au moins six minutes d'affilée. Lorsque l'audience totale d'un programme télévisé est exprimé en pourcentage du nombre de foyers, ce pourcentage est appelé taux d'écoute (38) du programme.

4. L'Audience moyenne : Le nombre de foyers écoutant un programme donné durant une minute moyenne de ce programme. Du fait que certains téléspectateurs changent de chaîne en cours de programme, l'audience totale est toujours supérieure à l'audience moyenne. Les audiences moyennes sont aussi exprimées sous forme de taux d'écoute et sont disponibles quart d'heure par quart d'heure.

5. La Part d'Audience : Il s'agit du pourcentage de foyers écoutant un programme donné, parmi tous les foyers à l'écoute d'un programme quelconque à un moment donné.

L'exemple suivant illustre le calcul de ces cinq index.

Cinq Mesures de l'Audience de la Télévision

Foyers	Situation	Programmes 1	2	3	Période Considérée
1	Pas de récepteur TV	0	0	0	de 7h. à 7h.15mn.
2	Récepteur arrêté	0	0	0	couverture = 9/10
3	Branché 1ère chaîne	1	0	0	Foyers récepteurs = 7/10
4	Branché 2ème chaîne	0	1	0	
5	Branché 3ème chaîne	0	0	1	Taux d'écoute
6	Branché 1ère chaîne	1	0	0	1ère chaîne = 3/10
7	Branché 2ème chaîne	0	1	0	2ème chaîne = 2/10
8	Branché 3ème chaîne	0	0	1	3ème chaîne = 2/10
9	Branché 1ère chaîne	1	0	0	
10	Récepteur arrêté	0	0	0	Part d'audience 1ère chaîne = 3/7 2ème chaîne = 2/7 3ème chaîne = 2/7

Pour chacune de ces cinq mesures, Nielsen fournit la partition selon la taille du comté (A,B,C,D), la taille du foyer (1-2, 3-4, 5+), l'âge du chef du ménage (35-, 34- 49,50+), le revenu du foyer (-$ 5 000, $ 5 000-10 000, $ 10 000+) et le niveau d'éducation (primaire, secondaire, supérieur). De plus Nielsen fournit aussi la répartition par âge et par sexe des audiences des programmes télévisés selon la grille suivante :

a) L'ensemble des femmes âgées de 18 ans ou plus est subdivisé en : 18-39, 18-49, 25-49, 25-64, 50 ans ou plus.

b) l'ensemble des hommes âgés de 18 ans ou plus est subdivisé en : 18-39, 18-49, 25-49, 25-64

c) l'ensemble des adolescents âgés de 12 à 17 ans est subdivisé en filles et garçons.

En outre, la technique du journal d'écoute permet d'obtenir des mesures d'audience cumulées qui seront décrites plus loin. Pour ses 230 zônes géographiques, Nielsen utilise des échantillons de ménage tirés au hasard, dont la taille va de 2200 ménages pour New-York jusqu'à 500 ménages pour les zônes géographiques les moins peuplées. L'ensemble de ces échantillons représente un total de 90 000 ménages. Pour les zônes géographiques les plus peuplées, les

journaux d'écoute sont distribués huit fois par an et chacun d'eux couvre une période de quatre semaines. Les zones les moins peuplées ne sont interrogées que trois fois par an. Pour les mesures nationales et non plus locales, Nielsen utilise un panel qui comprend 1200 ménages renouvelé pour un quart tous les ans.

Des données similaires sont collectées par American Research Bureau (39) le principal concurrent de Nielsen, qui utilise une méthode identique, à une exception près. Ses journaux d'écoute sont du type ouvert, c'est-à-dire qu'ils confient à l'enquêté le soin d'enregistrer l'heure du début et de la fin de l'écoute télévisée, tandis que Nielsen utilise des journaux d'écoute fermés où les heures d'écoute, déjà imprimées, n'ont plus qu'à être cochées.

Les deux autres sources de données concernant l'audience des programmes télévisés sont Simmons et TGI. Ce dernier, contrairement aux trois autres, n'utilise pas la technique des journaux d'écoute, mais fait appel à la mémoire des enquêtés. Ceux-ci doivent remplir un questionnaire postal au sujet des programmes regardés le jour précédent pour les programmes journaliers, ou la semaine précédente pour les programmes hebdomadaires. L'audience des diverses heures de la journée est obtenue en demandant aux enquêtés d'indiquer la part de leur temps consacré à la télévision pour chacune des périodes de la journée. Pour les retransmissions sportives, TGI mesure chaque sport séparément, en demandant aux enquêtés d'indiquer leur fréquence d'écoute de chacun des sports. Enfin, les nouveaux programmes qui sont lancés à la rentrée d'automne font l'objet d'un traitement différent. Pour ceux-ci, TGI utilise un journal d'écoute d'une durée d'une semaine appelé media log. Toutefois, le taux de retour de ces media logs est très faible puisqu'un peu plus de la moitié seulement sont retournés correctement remplis.

W.R. Simmons procède différemment. Les enquêteurs se rendent au domicile d'un échantillon de ménages désigné au hasard et sollicitent la coopération des enquêtés en vu de remplir un livre d'écoute pendant une période de deux semaines. Tous ceux qui acceptent sont familiarisés avec le livre d'écoute et l'enquêteur leur explique comment le remplir. Ils sont ensuite appelés par téléphone quelques jours après afin de détecter tout problème. A la fin de la première semaine, l'enquêteur vient chercher à domicile la première partie du journal d'écoute et contrôle si il a été correctement rempli. La deuxième partie du journal d'écoute est renvoyée par la poste.

Les limitations et les désavantages partagés par les quatre services de recherches sont les suivants :

A - Problèmes d'échantillonage :

Les échantillons utilisés par Nielsen et ARB sont tirés sur la base des annuaires téléphoniques et, par conséquent, éliminent les foyers ne possédant pas le téléphone. Les échantillons utilisés par TGI et Simmons sont construits sur la base du recensement de 1970 et, par conséquent, ne sont pas parfaitement à jour.

B - Problèmes de non réponse :

Les personnes ou les ménages qui refusent de coopérer ou qui retournent des questionnaires inexploitables peuvent posséder des habitudes d'écoute différentes de celles des personnes ayant répondu. Leur remplacement par d'autres individus, même tirés au hasard, modifie la représentativité théorique de l'échantillon puisque ceux qui n'ont pas répondu ont désormais une probabilité nulle d'être inclus.

C - Problèmes liés à la tenue des journaux d'écoute :

Les personnes qui acceptent de tenir un journal d'écoute ne s'en tiennent pas nécessairement à la lettre des instructions reçues. Une part inconnue mais certainement non négligeable des écoutes n'est pas enregistrée au moment où elles se produisent mais plus tard, sur la base de souvenirs, de vagues ouï-dires ou, pire encore, par simple extrapolation des comportements passés.

D - Problèmes liés au recours à la mémoire

Il est probablement plus aisé d'obtenir des réponses aux questions portant sur les écoutes de la veille ou de la semaine précédente, que de convaincre l'enquêté à s'abstreindre à la tenue d'un journal d'écoute. Cependant, bien des personnes ne se souviendront pas avec exactitude des programmes qu'ils ont vu la semaine dernière.

En résumé, les deux principaux inconvénients du journal d'écoute sont un taux de refus important et un coût de placement élevé. Par contre, l'avantage essentiel du journal d'écoute est qu'il ne met pas trop à l'épreuve la mémoire de l'enquêté, du moins si celui-ci enregistre ses écoutes au moment où elles se produisent. Les principales différences méthodologiques séparant les quatre services de recherches sont résumés dans le tableau suivant .

La mesure des Audiences de la Télévision aux Etats-Unis

Service de Recherche	Nielsen	ARB	TGI	Simmons
1. Méthode utilisée	Journal fermé	Journal ouvert	Souvenir d'écoute	Journal fermé
2. Sélection des enquêtés	Panel national + 230 échantillons régionaux	échantillons régionaux	Echantillon national	Echantillon national
3. Taille de l'échantillon	Panel de 1200 foyers + échant. locaux de 2200 à 600 foyers	échantillons de 4000 à 400 foyers	18 300 individus	8 800 individus
4. Nombre de journaux ou de questionnaires	Panel = 1100 locaux = 300	Supérieur à 100 par région	10 700 individus	6 000 individus
5. Taux de réponses	56 %	55 %	58 %	68 %
6. Périodicité des mesures	8 fois par an	8 fois par an	1 fois par an (Oct. Nov.)	1 fois par an (Oct. Nov.)
7. Méthode de placement	Voie postale	Postale et téléphone	contact tél. puis voie postale	Enquêteur à domicile
8. Période couverte par le journal ou le questionnaire	4 semaines	1 semaine	la dernière semaine	2 semaines
9. Stimulant	50 ¢ par journal et entretien téléviseur pour les panélistes	50 ¢ par journal	Cadeaux	Cadeaux

Ces différences méthodologiques influent beaucoup sur les résultats obtenus. Pour s'en convaincre, il suffit d'examiner les deux tableaux suivants qui comparent les audiences estimées par TGI, Simmons et Nielsen pour plusieurs catégories de programmes diffusés en soirée par les trois principales chaînes américaines NBC, ABC et CBS.

Audiences Adultes de 11 Programmes Télévisés

diffusés en soirée - Automne 1971

Audiences estimées par			
Programmes	Simmons	Nielsen	Différence en pourcentage
Glen Campbell	16,116	15,520	3,8 %
CBS Evening News	13,715	10,700	28,2 %
Bold Ones	15,401	11,180	37,7 %
Flip Wilson	31,027	17,370	78,6 %
Medical Center	23,284	15,810	47,3 %
Ironside	16,400	16,650	-1,5 %
ABC News	4,661	4,930	-5,8 %
ABC Sunday Night Movie	19,278	13,090	47,8 %
Dean Martin	17,060	14,500	16,8 %
Marcus Welby	26,810	17,430	53,8 %
Laugh In	18,209	14,420	26,3 %

Audiences Adultes de 8 Types de Programmes Télévisés

diffusés en soirée - Automne 1972

Programmes	Audiences estimées par TGI	Simmons	Différence en pourcentage
Informations			
NBC (Chancellor)	15,297	8,852	72,8 %
CBS (Cronkite)	26,269	13,634	92,7 %
CBS (Mudd) Saturday	13,593	6,753	100,5 %
Comédie et Variété			
Flip Wilson (NBC)	34,361	21,404	60,5 %
Dean Martin (NBC)	18,949	12,237	54,8 %
Sonny & Cher (CBS)	28,563	14,734	33,8 %
Grand Film			
Monday Night Movie (NBC)	35,030	19,182	82,6 %
Sunday Night Movie (NBC)	25,968	29,371	-13,1 %
Friday Night Movie (CBS)	27,070	16,740	61,7 %
Séries Dramatiques			
Marcus Welby MD (ABC)	40,056	25,242	58,7 %
Medical Center (CBS)	24,802	20,216	22,7 %
Emergency (NBC)	26,316	11,689	125,1 %
Séries Policières			
Hawaii Five O (CBS)	35,481	21,945	61,7 %
Adam 12 (NBC)	30,624	16,344	87,4 %
The Rookies (ABC)	24,101	14,550	65,6 %
Séries Détectives			
Cannon (CBS)	31,106	22,117	40,6 %
Mannix (NBC)	27,283	14,480	88,4 %
Séries Comiques			
All in the Family (CBS)	46,326	36,789	25,9 %
Love Americain Style (ABC)	20,613	9,598	115,9 %
Flip Wilson (NBC)	34,361	21,409	60,5 %
Sports			
NFL Monday Night Football (ABC)	34,151	15,525	119,9 %

Ces comparaisons indiquent que les estimations de TGI sont très supérieures à celles de Simmons qui sont elles-mêmes supérieures à celles de Nielsen. Les audiences télévisées estimées par TGI sont parfois trois fois plus élevées que leur contrepartie chez Nielsen ! De telles différences mettent en question le recours à la mémoire des enquêtés en tant qu'instrument de mesure des comportements d'écoute télévisée. Les estimations d'ARB qui n'étaient pas disponibles pour notre comparaison sont très proches de celles de Nielsen. En fin de compte, les trois services de recherche d'audience qui utilisent la technique des journaux d'écoute, obtiennent des estimations, certes, différentes, mais l'importance de ces différencesn'est pas supérieure à celles observées dans le domaine des magazines. Les acheteurs d'espace télévisé ont montré une réticence compréhensible à utiliser les données fournies par TGI et par Simmons. Ils continuent à s'en tenir aux estimations plus anciennes et probablement plus fondées produites par Nielsen et ARB. Cependant, ces deux derniers ne mesurent pas l'audience des magazines et ceci crée un problème difficile à résoudre pour les annonceurs qui utilisent simultanément les magazines et la télévision, car on ne voit pas très bien sur quelle base fusionner des données non comparables. De plus, les estimations fournies par Nielsen et ARB sont mises à jour huit fois par an,tandis que celles de TGI et Simmons ne sont collectées qu'une fois par an et ne sont pas subdivisées en zones géographiques détaillées. Enfin, les estimations de TGI et de Simmons sont subdivisées selon de nombreux critères socio-démographiques et selon le taux d'usage d'un grand nombre de produits et de marques, ce qui n'est pas le cas des estimations de Nielsen et d'ARB.

En ce qui concerne l'écoute des stations de radio, les deux principales techniques de mesure sont le journal d'écoute et l'enquête téléphonique. Le journal d'écoute radio est très similaire au journal d'écoute télévision. Les deux journaux se partagent parfois une même page du livret d'écoute comme c'est le cas en France où la société Secodip mesure simultanément l'écoute radio et l'écoute télévisée auprès de ses 9000 panélistes. Les enquêtes téléphoniques sont de deux types: L'enquête en direct et l'enquête en différé (40). Dans la première on demande à l'enquêté le nom de la station qu'il était en train d'écouter au moment de l'appel téléphonique, dans la seconde on avertit l'enquêté qu'on le rappellera dans 24 heures pour lui demander ce qu'il a écouté. Dans l'enquête en différé, le premier appel permet de s'assurer de la participation de l'enquêté et d'effectuer par la suite des appels quotidiens au cours desquels les écoutes des divers membres de la famille pourront être enregistrés séparément. Les avantages et les inconvénients des deux types d'enquête téléphonique sont les suivants :

- l'enquête téléphonique permet d'obtenir un taux de réponses supérieur au journal d'écoute, spécialement pour l'enquête en différé, et tout spécialement parmi les groupes ethniques peu alphabétisés.

- l'enquête en différé est susceptible de modifier les comportements d'écoute du fait que l'enquêté sait qu'il sera interrogé le lendemain.

- L'enquête en direct ne permet pas de mesurer les audiences cumulées puisque chaque individu n'est interrogé qu'une fois.

- Les personnes équipées du téléphone ne constituent pas un échantillon représentatif de l'ensemble de la population, spécialement en France.

L'American Research Bureau, qui est le chef de file dans le domaine des mesures radiophoniques, mesure l'audience des stations de radio pour 163 zones géographiques, sur une base annuelle ou trimestrielle, selon l'importance des régions. Pour chacune des régions, la taille de l'échantillon varie entre 500 et 2000 journaux d'écoute selon l'importance de la population. Les données fournies comprennent l'audience moyenne, quart d'heure par quart d'heure, l'audience cumulée sur quatre semaines et la part d'audience de chacune des stations concurrentes. Ces données sont subdivisées en écoute à domicile et écoute hors du domicile, et en auditeurs métropolitains et auditeurs non métropolitains. De plus, diverses répartitions de l'audience selon des critères démographiques sont disponibles ainsi que des analyses de l'évolution des audiences en cours d'année. Selon ARB, 85% des ménages contactés acceptent de participer et 65% des journaux d'écoute retournés sont utilisables, ce qui donne un taux de réponses globales de 52%.

La seconde source de données radiophoniques est la société Pulse (41) qui couvre 150 zones géographiques de façon régulière et 100 autres zones sur demande spéciale. Pulse utilise la technique des enquêtes personnelles à domicile et mesure le souvenir laissé par les écoutes. On demande aux enquêtés d'indiquer la station écoutée la veille et les heures d'écoute. Les réponses sont utilisées pour calculer l'audience moyenne, quart d'heure par quart d'heure, de chaque station. Cependant, afin d'estimer les audiences cumulées, on demande aussi à chaque enquêté d'indiquer la ou les stations écoutées, à n'importe quel moment de la journée au cours de la semaine passée. Les réponses à cette question sont nécessairement imprécises à cause des distorsions mémorielles.

La troisième source d'informations radiophoniques est la société C.E. Hooper (42) qui couvre 100 régions géographiques en utilisant la méthode des enquêtes téléphoniques. En dehors de ces trois sources principales, trois autres sociétés sont spécialisées dans les études"à la carte" : Market Trends (43) qui couvre la région des grands lacs, Trendex (44) qui couvre la région new-yorkaise et The Source (45) qui couvre la côte ouest.

Simmons et TGI ne mesurent pas l'audience des stations de radio individuelles mais l'audience moyenne du medium radiophonique dans son ensemble. Les données collectées sont similaires à celles déjà décrites pour les quotidiens. On dispose ainsi pour chaque période de la journée de l'audience moyenne selon le sexe, l'âge, le revenu, l'éducation et l'importance du comté. TGI distingue entre l'audience moyenne des jours de semaine et l'audience des week-end. Il fournit aussi la distribution des audiences selon treize types de programmes : le rock, le hit parade, la musique populaire, la musique de toujours, la musique instrumentale, la musique classique, la musique folklorique, les shows avec célébrités, les informations, la musique noire, le sport, les émissions destinées aux agriculteurs et les émissions religieuses.

Le tableau suivant résume les différences méthodologiques existant entre les six services de recherches spécialisées dans les écoutes radiophoniques.

La Mesure des Audiences Radiophonique aux Etats-Unis

Service de Recherche	Arbitron	Hooper	Market Trends	Pulse	The Source	Trendex
1. Méthode utilisée	Journal d'écoute Téléphone en différé	Téléphone en direct	Téléphone en différé	Enquêteur à domicile	Enquêteur et journal d'écoute	Téléphone en différé
2. Sélection des enquêtés	162 régions	100 régions	5 régions	150 régions	Los Angeles	New-York
3. Taille de l'échantillon	500 à 2500	2200	1800 à 3500	500 à 5500	3000	2300
4. Taux de réponses	52 %	--	--	60 %	--	51 %
5. Périodicité	4 fois par an	sur demande	4 à 8 fois par an selon la région	1 à 6 fois par an selon la région	2 fois par an	3 fois par an
6. Données collectées	Audience moyenne par 1/4 d'heure à domicile et hors du domicile Urbaines non urbaines Répartitions démographiques tendances annuelles	Parts d'audiences et répartitions démographiques	Audience moyenne par 1/4 d'heure parts d'audience Audiences cumulées Répartitions démographiques	Audience moyenne par 1/4 d'heure Audiences hebdomadaires cumulées Répartitions démographiques	Audience moyenne par 1/4 d'heure Audiences hebdomadaires cumulées Répartition par produit	Audience moyenne par 1/4 d'heure Audiences cumulées Parts d'audience Répartitions démographiques

Compte tenu des différences méthodologiques que nous venons d'inventorier, il n'est guère surprenant d'observer que les audiences des stations de radio varient selon qui les mesure. Le tableau suivant, extrait des documents commerciaux publiés par ARB (46), montre que les taux d'écoute métropolitaine estimés par Pulse sont supérieurs de 100% en moyenne à ceux d'ARB sur la base de 15 villes. Cette information conforte l'observation similaire que nous avons déjà faite en ce qui concerne la télévision. Les estimations basées uniquement sur le souvenir conduisent à des audiences plus élevées que les estimations basées sur la tenue d'un journal d'écoute.

Comparaison des Audiences Radiophoniques

mesurées par Pulse et par ARB

(Taux d'écoute horaire moyen de 6 heures à 24 heures parmi les personnes âgées de 12 ans ou plus)

Ville	ARB	Pulse	Différence
Akron	40	63	+ 23
Boston	33	42	+ 9
Buffalo	49	58	+ 9
Chicago	27	53	+ 26
Cincinnati	25	71	+ 46
Dayton	41	82	+ 41
Denver	24	115	+ 91
Fresno	43	86	+ 43
Kansas City	39	56	+ 17
Menphis	42	106	+ 64
Miami	45	102	+ 57
Omaha	39	47	+ 8
Philadelphia	40	50	+ 10
Washington, D.C.	34	74	+ 40
TOTAL	574	1093	+ 519

Nous avons terminé notre étude critique des différentes méthodes qui permettent d'estimer l'exposition aux divers supports publicitaires. Nous sommes maintenant en mesure de préciser le contenu du concept d'audience.

4 - L'Audience

Ainsi que nous venons de le voir, le concept d'exposition se définit par l'existence d'une opportunité de voir ou d'entendre un support publicitaire. La classe d'individus pour lesquels cette relation est vérifiée constitue l'audience du support. Cette classe d'individus peut être subdivisée en sous-classes par la spécification d'un certain nombre de caractères devant être possédés par les individus appartenant à chacune des sous-classes. Ainsi, les hommes ayant de 24 à 45 ans, gagnant $ 15.000 ou plus, vivant dans le comté de Westchester et ayant eu au moins une opportunité de lire un numéro donné de Newsweek, constitue une audience particulière de ce magazine.

Puisque les audiences sont par essence mouvantes et dynamiques, leur définition implique la spécification d'une période de référence et d'un lieu de référence. En fin de compte la composition d'une audience a tout autant d'importance que son volume global. Il est bien difficile de donner une liste complète de tous les critères susceptibles d'être utilisés pour décomposer une audience. Nous nous contenterons de recenser dans la table suivante les critères les plus communément employés, en les regroupant en quatre catégories :les critères liés à l'individu (qui ?), les critères liés à l'espace (où), les critères liés au temps (quand ?) et les critères liés à l'activité des individus (quoi ?).

LES CRITERES USUELS DE DECOMPOSITION DES AUDIENCES

I - LA DEFINITION DES INDIVIDUS

A - Critères Socio-Démographiques

1 - Sexe
2 - Age
3 - Revenu
4 - Profession
5 - Education
6 - Etat-Civil
7 - Taille de la famille
8 - Race
9 - Classe Sociale
10 - Religion
etc ..

B - Critères Psychographiques

1 - Auto-Portrait (Self-Concept) : La manière dont une personne se décrit elle-même en fonction d'une liste d'adjectifs tels que : affectueux, amical, timide, brave, tolérant, créatif, dominateur, efficace, égocentrique, franc, drôle, réservé, intelligent, raffiné, sociable, tendu, têtu, etc ..

2 - Comportement d'achat (Buying Style) : La manière dont une personne décrit ses propres habitudes d'après une liste d'adjectifs tels que : Fidèle à une marque, prudent, conformiste, économe, recherchant la nouveauté, impulsif, planifié, etc. . . .

II - LA DEFINITION DU LIEU

A Lieu Géographique

1 - Régions de programme
2 - Départements
3 - Villes
4 - Communes
etc. . . .

B - Lieu Sociologique

1 - Habitat urbain
2 - Habitat suburbain
3 - Habitat campagnard
4 - Résidence principale vs. résidence secondaire
5 - Exposition chez soi vs. Exposition sur les lieux de travail, vs. exposition durant le trajet jusqu'au lieu de travail, etc. . . .

III - LA DEFINITION DU MOMENT

A - Le Temps Historique

1 - L'année
2 - La saison
3 - Le mois
4 - La semaine
5 - Le jour
6 - L'heure

B - Le Temps Psychologique

1 - Jour de travail vs. jour férié
2 - Le matin, le midi, l'après-midi, le soir
3 - Pendant le repas, entre les repas
4 - Temps de repos vs. temps actif

IV - LA DEFINITION DE L'ACTIVITE

A - Activités Liées au Support

1 - Degré d'attention (vs. distraction)
2 - Support reçu par abonnement vs. achat impulsif
3 - Support acheté vs. support emprunté
4 - Degré de fréquentation du support
5 - Degré d'appréciation du support (échelle de satisfaction)
etc. . . .

B - Activités Liées au Produit

1 - Degré de consommation du produit
2 - Degré de consommation de la marque
3 - Degré de fidélité à la marque
4 - Fréquence d'achat
5 - Date du dernier achat
6 - Degré de satisfaction vis à vis de la marque
7 - Intentions d'achat

En résumé, le spécialiste de la planification des supports dispose, de nos jours, d'une masse d'informations qui lui permettent d'évaluer l'aptitude de chaque support à couvrir un groupe de consommateurs cible préalablement défini. On peut considérer que les dix dernières années ont été le témoin d'un bond énorme dans la quantité de données d'audience disponibles. Cet accroissement quantitatif s'accompagne aussi d'un saut qualitatif. En effet, jusqu'à une époque récente, les audiences étaient décomposées, critère par critère, d'une façon séquentielle. Il était très rare de pouvoir utiliser plus de deux critères simultanément pour définir une audience. Aujourd'hui, grâce aux mémoires magnétiques à accès rapide, il est possible de croiser ensemble un grand nombre de critères et de parvenir ainsi à une définition multicritère de la cible extrêmement précise. Par exemple, on peut connaître le nombre de lecteurs de Newsweek possédant une voiture de marque A, achetée il y a trois ans et qui ont exprimé l'intention d'acheter une nouvelle voiture dans les six mois à venir. Aucune de ces classifications, prise séparément, ne serait très utile. La possibilité de les combiner simultanément permet de décider s'il convient de placer la publicité dans Newsweek plutôt que dans Time. Néanmoins, plus la cible est définie avec précision, plus la base de sondage devient étroite et plus l'incertitude associée à l'estimation de l'audience s'accroît.

Cependant, la comparaison des audiences moyennes de deux supports pour une cible donnée n'est qu'une partie du problème. La plupart des campagnes publicitaires comportent plusieurs insertions dans un même support et font appel à plusieurs supports simultanément. Afin de mieux cerner les complexités supplémentaires qui en résultent, il nous faut maintenant définir plus précisément les concepts de couverture nette, d'accumulation et de duplication qui nous permettront de mesurer les dimensions dynamiques des audiences.

5. La Couverture nette

Avant d'examiner le concept de couverture nette il nous faut définir l'objet auquel il s'applique : le plan publicitaire. Un plan publicitaire est une liste d'insertions publicitaires datées, à paraître dans une liste spécifiée de supports publicitaires. Ce plan possède donc deux dimensions : le temps et les supports . La table suivante illustre une manière pratique de décrire un plan publicitaire.

Dimension des supports	Dimension temporelle			
	1ère semaine	2ème semaine	3ème semaine	4ème semaine
1 - Télévision				
Today shows(NBC)	X	-	X	-
CBS Evening News	X	X	X	X
2 - Magazines				
Newsweek	X	X	X	X
Time	-	X	-	X

Dans ce simple exemple, le plan ne comprend que deux media, la télévision et les magazines, avec deux supports pour chacun d'eux, et un total de 12 insertions à raison de 3 par semaine pour une période de 4 semaines. Un plan publicitaire réel comprend davantage d'informations telles que le format de chaque insertion, sa durée, les émissions dans lesquelles elles s'insèrent s'il s'agit d'un support radiodiffusé, ou encore, la taille, la couleur et la position des messages s'il s'agit d'un support imprimé. Tout cela est complété par une description détaillée de chacun des messages utilisés ainsi que par la définition de la cible visée. Toutes ces informations sont nécessaires à une définition précise de la couverture nette. La couverture nette d'un plan publicitaire est égale au nombre d'individus cibles ayant été exposés au moins une fois à l'un des supports utilisés pendant une période de temps déterminée.

Cette définition mérite quelques remarques. Tout d'abord,il est incorrect de parler de la couverture nette d'une parution d'un support, le terme d'audience est plus approprié. En second lieu la couverture nette est un concept qui a une dimension temporelle. L'emploi d'une période de référence de quatre semaines tend à se généraliser parmi les praticiens. En conséquence, chaque fois qu'il n'est pas fait mention d'une période de référence précise, il est entendu qu'il s'agit d'une couverture nette de quatre semaines. En troisième lieu, la couverture nette se rapporte en général à un plan d'insertions dans plusieurs supports. Le détail de ce plan doit donc être connu sous peine de vider le concept de tout son sens. Enfin, la couverture nette se rapporte à une population cible qui doit avoir été définie préalablement. Il convient aussi de ne pas confondre le concept de couverture nette avec le concept d'impact. La couverture nette mesure le nombre d'individus cibles ayant eu au moins une opportunité d'entrer en contact avec au moins un des supports utilisés, tandis que l'impact

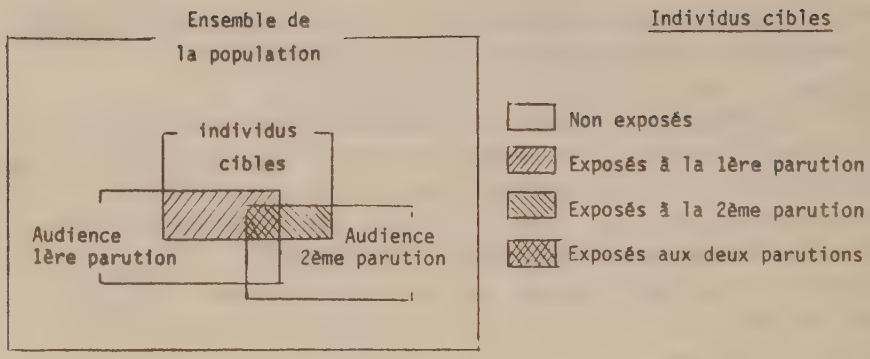
mesure le nombre d'individus cibles ayant eu au moins une opportunité d'entrer en contact avec l'un des messages de la campagne publicitaire.

La définition précédente indique clairement que, quelle que soit la cible visée, la couverture nette dépend à la fois du nombre de supports utilisés, du nombre d'insertions effectuées dans chacune d'eux et de la durée de la campagne. Par conséquent, la question clef est la suivante : quel accroissement de la couverture nette peut-on attendre de l'adjonction d'une nouvelle insertion, soit dans l'un des supports déjà utilisé, soit dans un nouveau support ? La réponse à cette question réside dans les deux concepts d'accumulation et de duplication d'audience.

6. L'Accumulation des Audiences

Tous les supports, ou du moins ceux qui ne sont pas détruits du fait même de leur rencontre avec leur première audience (47), voient leur audience croître au fil du temps. La vitesse d'accumulation est une caractéristique fondamentale de chaque support. Il convient de distinguer deux types d'accumulation. L'accumulation résultant des parutions successives d'un même support et l'accumulation résultant des prises en main successives d'un même support par différents individus. Le premier phénomène est absolument général et vaut pour tous les supports. Le second phénomène ne s'applique qu'aux supports imprimés caractérisés par une certaine durée de vie active pendant laquelle ils sont susceptibles d'entrer en contact avec d'autres lecteurs que le lecteur initial. On utilisera le terme de lecteurs secondaires pour désigner l'ensemble des individus qui entrent en contact avec un journal ou un magazine, sans l'avoir acheté ni s'y être abonné et on les incluera dans l'audience du support. Le terme d'accumulation sera donc réservé à la description du premier phénomène et nous définirons l'audience cumulée de N parutions d'un support par le nombre d'individus exposés à l'une quelconque de ces parutions quelles que soient les circonstances de l'exposition.

Prenons le cas de deux parutions d'un magazine. Parmi les individus cibles certains feront partie de l'audience de la première parution mais non de la seconde, certains autres feront partie de l'audience de la seconde parution mais non de la première, d'autres enfin feront partie de l'audience des deux parutions et le reste des individus cibles ne sera ni dans l'audience de la première parution ni dans celle de la seconde. Le diagramme suivant illustre la situation .



L'audience cumulée de deux parutions de ce magazine est égale au nombre d'individus cibles exposés à une parution au moins. Les individus exposés aux deux parutions constituent l'audience dupliquée. Dans le cas de trois parutions, les individus exposés trois fois constituent l'audience tripliquée et ainsi de suite. Les individus qui n'ont été exposés qu'à une parution et à une seule constituent l'audience exclusive de cette parution. Il est clair que, toutes choses égales par ailleurs, plus l'audience dupliquée est forte, plus l'audience cumulée sera faible. Réciproquement, plus l'audience est volatile, plus elle se renouvelle d'une parution à une autre, plus large sera la couverture nette cumulée et plus fort sera le taux d'accumulation. En d'autres termes, les supports jouissant d'un fort pourcentage de lecteurs ou d'auditeurs fidèles accumuleront moins d'audience que les supports où les lecteurs ou auditeurs occasionnels prédominent. Une mesure utile de la capacité d'accumulation d'un support est donnée par le taux de rotation des audiences t.

$$t = \frac{\text{Accroissement d'audience dû à la 2ème parution}}{\text{Audience d'une parution}}$$

Moyennant une hypothèse particulière concernant la distribution des occasions de voir nous démontrerons dans le chapitre 5 que la seule connaissance du taux de rotation et de l'audience d'une parution permet de calculer l'audience cumulée d'un nombre quelconque de parutions.

Ainsi qu'on peut le supposer, il n'y a pas davantage de consensus en ce qui concerne la mesure des audiences cumulées qu'en ce qui concerne la mesure de l'audience moyenne d'une parution. Théoriquement, il suffirait de recenser,

en fin de période, les individus déclarant avoir été exposés à l'une quelconque des parutions de la période étudiée. Néanmoins, fort peu d'individus sont capables de se souvenir d'une lecture ou d'une écoute intervenue plusieurs semaines auparavant. La difficulté peut être surmontée de deux manières. On peut, soit interroger la même personne à intervalles réguliers ou lui faire tenir un journal d'écoute ou de lecture, soit extrapoler sur la base des habitudes de lecture et d'écoute. La première méthode fournit une solution directe mais coûteuse. De plus, la répétition des questions risque de modifier les comportements mesurés. La seconde méthode, bien qu'indirecte, possède l'avantage de l'économie puisqu'elle ne nécessite qu'une seule observation. Néanmoins, elle doit être préalablement validée.

61. Les Mesures directes

Aux Etats-Unis il n'existe pas de panel pour les supports imprimés. Simmons utilise la méthode de la double enquête qui consiste à interroger deux fois la même personne, à 4 ou 5 semaines d'intervalle, sur ses lectures reconnues de deux numéros différents. Simmons a repris en celà la technique mise au point par Politz qui utilisa tout d'abord, en 1950 et 1952 jusqu'à 13 enquêtes successives, puis en 1958, six enquêtes successives. L'analyse des courbes d'accumulation recueillies par ces premières études permit de conclure que la connaissance de l'audience cumulée sur deux numéros suffisait pour prédire avec une bonne précision l'accumulation des numéros suivants. La nécessité d'étudier un nombre de plus en plus élevé de magazines s'est donc accompagnée d'une diminution du nombre de numéros couverts par l'enquête. En 1965, Simmons (29) crut pouvoir se dispenser de la double enquête en interrogeant, en une seule et même occasion, sur les lectures de deux numéros différents. On s'aperçut alors que l'enquête avait tendance à déclarer avoir lu soit les deux numéros, soit aucun. Cette tendance reflète le besoin naturel, éprouvé par tout enquêté, d'apparaître cohérent dans ses réponses. Ainsi, un lecteur occasionnel ayant déclaré lire l'un des numéros aura tendance à déclarer avoir lu aussi le second, même s'il n'en est rien. Il en résulte une réduction du taux d'accumulation d'audience et une surestimation du nombre de lecteurs fidèles. Par contre, si on procède par double enquête séparée par un mois d'intervalle, il est fort probable que le lecteur occasionnel perde le souvenir de sa réponse à l'enquête précédente, ce qui met fin au biais précédemment signalé.

Néanmois, la technique de la double enquête se heurte au fait qu'un pourcentage non négligeable d'enquêtés ne peuvent être ré-interrogés du fait

de leur absence ou de leur refus de participer à nouveau. Toutefois, ce problème se pose également au niveau des panels et rien ne prouve que les réfractaires aient un comportement de lecture nettement différent de celui de la population. En France, il existe deux panels média-produits, gérés par Secodip. Ces panels utilisent la technique des livres de lecture et d'écoute et permettent de calculer directement les audiences cumulées de 21 hebdomadaires et de 20 mensuels. Toutefois, leurs résultats sont limités aux audiences des ménagères, uniquement. De plus, la période de collecte des données ne s'étend que sur un trimestre, d'Avril à Juin, ce qui limite le cumul des audiences à 12 semaines pour les hebdomadaires et à 3 mois pour les mensuels. Pour des périodes plus longues il devient nécessaire, ici aussi, de recourir à des méthodes d'extrapolation.

62. Les Mesures indirectes

T.G.I. aux Etat-Unis et le CESP en France se limitent à une enquête ponctuelle. Ils doivent donc recourir à une estimation indirecte basée sur les habitudes de lecture déclarées. Les réponses à la question d'habitudes sont étalonnées à l'aide des lectures de la dernière période comme nous l'avons vu précédemment. Toutefois, cela ne garantit pas une estimation correcte des audiences cumulées car, en dehors des biais mémoriels et du biais dû aux lectures répétées, la méthode est exposée à 4 sources supplémentaires d'erreur que nous étudierons en détail au chapitre 5.

- L'effet de découpage qui tient au remplacement des probabilités individuelles, par essences continues, par des probabilités discrètes et en nombre fini affectées à chacune des classes d'habitude.

- L'effet d'agrégation qui tient à l'hétérogénéitée des probabilités individuelles d'exposition à l'intérieur de chaque classe d'habitude.

- L'effet sémantique qui tient aux différences d'interprétation, d'un individu à un autre, de la signification des échelons verbaux utilisés pour décrire les habitudes d'exposition.

- L'effet d'instabilité qui tient au caractère mouvant des habitudes d'exposition et qui conduit certains individus à changer de classe d'habitude pendant la période étudiée.

Nous montrerons, dans le prochain chapitre, que les deux premiers effets contribuent à la sous-estimation de l'audience cumulée, tandis que les deux derniers effets agissent en sens contraire. Avant de clore cette discussion consacrée à la mesure des audiences cumulées, il est intéressant de

procéder à la comparaison des audiences mesurées directement par Simmons selon les méthodes de la double enquête avec des audiences extrapolées à partir des habitudes de lecture par T.G.I.

Comparaisons des Audiences Cumulées de 9 Magazines Américains en 1973

	Simmons (mesure directe par double enquête)			T.G.I. (mesure indirecte par extrapolation des habitudes de lectures)		
Magazines	1 numéro	2 numéros	taux d'accumulation	1 numéro	2 numéros	taux d'accumulation
Better Homes & Gardens	23020	31358	0,362	29319	38165	0,302
Family Circle	15674	21298	0,359	23676	30594	0,292
Good Housekeeping	16114	23150	0,437	22794	30268	0,328
Ladies'Home Journal	16679	23220	0,392	18529	23887	0,289
Newsweek	13985	20098	0,434	16229	22826	0,406
Time	20832	28378	0,362	15666	21365	0,364
U.S News & World Report	9025	12103	0,341	6966	9605	0,378
Reader's Digest	42108	52016	0,235	39431	49466	0,254
TV Guide	38937	51900	0,331	50617	59699	0,179

Cette comparaison montre que la mesure indirecte, basée sur les habitudes de lecture, conduit à des audiences supérieures à celles obtenues par la mesure directe sous forme de double enquête. Cependant les magazines Time, U.S. News & World Report et Reader's Digest font exception. Il convient de de noter que l'écart entre les deux méthodes est moindre au niveau de l'audience cumulée de deux parutions qu'au niveau de l'audience d'une parution. La situation est renversée en ce qui concerne le taux d'accumulation qui est plus élevé par la méthode directe que par la méthode indirecte, avec les trois mêmes exceptions.

7. La Duplication des Audiences

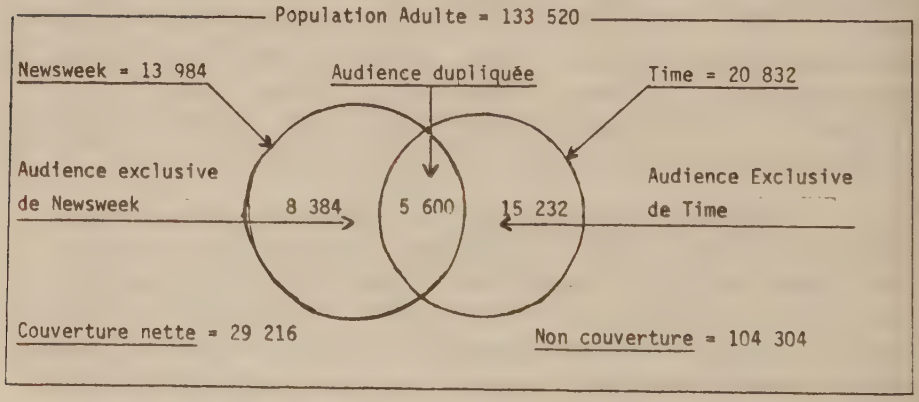
Tandis que le concept d'accumulation décrit la dynamique des audiences lorsque plusieurs parutions d'un même support sont considérées, le concept de

duplication décrit la dynamique des audiences résultant de la considération de plusieurs supports. L'audience dupliquée de m supports est égale au nombre d'individus exposés à chacun d'eux.

Dans le cas de deux supports, quatre segments d'audience peuvent être distingués. Les individus exposés au premier support mais non au second. Les individus exposés au second support mais non au premier. Les individus exposés aux deux supports. Les individus non exposés. Les deux premiers segments constituent l'audience exclusive de chacun des deux supports. Le troisième segment constitue l'audience dupliquée de la paire de supports. Les trois premiers segments constituent la couverture nette des deux supports, parfois appelée audience non dupliquée. Le dernier segment constitue la non couverture des deux supports. Le diagramme suivant illustre ces distinctions pour les magazines Time et Newsweek d'après les estimations de Simmons pour l'année 1973 et pour l'ensemble de la population adulte américaine.

Duplication de Time et Newsweek

(1973 , milliers d'adultes)



La couverture nette des deux magazines s'obtient en additionnant leurs audiences exclusives avec l'audience dupliquée soit :

$$8\ 384 + 15\ 232 + 5\ 600 = 29\ 216$$

Elle peut aussi être calculée directement à partir des audiences de chaque magazine en soustrayant l'audience dupliquée de l'audience totale ou audience brute (48) :

Couverture nette = Audience brute - Audience dupliquée

Couverture nette = 13984 + 20832 - 5600 = 29216

Les définitions que nous venons de donner peuvent être complétées par les deux remarques suivantes :

- Du point de vue éthymologique, le terme de duplication devrait être réservé aux plans publicitaires ne comportant que deux supports. Dans le cas de 3 supports on devrait employer le terme triplication, pour 4 supports le terme quadruplication et ainsi de suite. Toutefois, et suivant en cela l'usage en cours dans les milieux publicitaires, nous étendrons le sens du mot duplication à tous les recouvrements d'ordre supérieur à deux, chaque fois qu'aucune confusion n'est possible.

- Le terme d'audience dupliquée est parfois utilisé pour désigner les individus exposés à deux parutions d'un même support. Bien que l'analogie soit évidente nous préférons utiliser le terme d'audience fidèle pour éviter toute confusion. Nous réservons l'usage du terme duplication aux plans publicitaires comprenant plusieurs supports avec une seule insertion par support. Quand un plan publicitaire comprend plusieurs supports avec plusieurs insertions dans chacun d'eux nous utiliserons le concept de matrice d'exposition.

8. La Matrice d'Exposition

Avant de donner une définition formelle du concept, il est utile de considérer un exemple. Le diagramme d'exposition aux deux magazines Time et Newsweek présenté à la page précédente peut être converti en la matrice d'exposition suivante.

Matrice d'Exposition Time - Newsweek

(1 parution de Time + 1 parution de Newsweek)

		TIME		Total
		Non Exposés	Exposés	Newsweek
NEWSWEEK	Non Exposés	104 304	15 232	119 536
	Exposés	8 384	5 600	13 984
	Total Time	112 688	20 832	133 520

Les lignes de la matrice d'exposition décrivent l'exposition au magazine Time en fonction de l'exposition au magazine Newsweek. Les colonnes de la matrice décrivent l'exposition à Newsweek en fonction de l'exposition de Time. La matrice d'exposition peut être étendue aisément à la description simultanée de l'accumulation et de la duplication résultant de deux parutions de Time et de deux parutions de Newsweek.

Matrice d'Exposition Time-- Newsweek

(2 parutions de Time + 2 parutions de Newsweek)

		Expositions à TIME			Total
		Zéro	Une	Deux	Newsweek
Expositions	Zéro	95 355	9 591	8 525	113 471
à	Une	5 978	3 728	2 421	12 127
Newsweek	Deux	3 809	1 772	2 340	7 921
	Total Time	105 142	15 091	13 286	133 520

Les totaux en ligne et en colonne de la matrice d'exposition décrivent la manière dont les audiences de chacun des deux magazines s'accumulent, tandis que les éléments de la matrice décrivent la manière dont les deux audiences se recouvrent mutuellement. La couverture nette du plan s'obtient immédiatement en retranchant de la population totale l'élément zéro-zéro de la matrice. Dans l'exemple ci-dessus on obtient une couverture nette de :

$$133\ 520 - 95\ 355 = 38\ 165$$

Quand trois supports sont considérés, nous obtenons une matrice tri-dimensionnelle et, plus généralement, avec m supports et n insertions dans chaque support, nous obtenons une matrice à m dimensions comportant $(n + 1)^m$ éléments. On conçoit aisément que, bien qu'il soit possible de tabuler les résultats d'enquête pour obtenir de telles matrices, cela requiert des calculs fort coûteux à partir de la bande magnétique où sont consignés, individu par individu, les résultats de l'enquête d'audience. En fait, le nombre d'éléments de la matrice d'exposition est une fonction exponentielle du nombre de supports. Par exemple, un plan faisant appel à quatre supports avec deux insertions dans chacun d'eux, donne naissance à une matrice d'exposition comportant 81 éléments. On atteint

ainsi très vite un point où le nombre d'enquêtés correspondant à chaque case de la matrice est si faible qu'il enlève toute fiabilité aux résultats. De plus, dès que le plan publicitaire prévoit plus de deux insertions par support, aucune donnée d'enquête n'est disponible. On comprend donc la nécessité où l'on se trouve de développer des modèles d'exposition permettant d'estimer tous les éléments de la matrice d'exposition à partir de la simple connaissance des audiences, des duplications par paire et des audiences cumulées à deux parutions, qui sont les seules données disponibles.

Bien que la matrice d'exposition procure une information exhaustive sur la structure des expositions engendrées par un plan publicitaire, cette information doit être réduite à une forme simple plus directement interprétable. Ceci est obtenu en recourant au concept de distribution des occasions de voir.

9. La Distribution des Occasions de Voir

La distribution des occasions de voir (ou d'entendre) énumère le nombre d'individus cibles ayant été exposés à zéro, une, deux, ...etc, parmi les opportunités de voir (ou d'entendre) présentées par le plan publicitaire. Le nombre de ces opportunités est égal à la somme du nombre d'insertions effectuées dans chacun des supports utilisés. Par exemple, dans le cas de deux insertions dans Time et Newsweek, certains individus auront été exposés aux 4 insertions, certains à 3, 2 ou 1 insertion alors que d'autres n'auront pas été exposés du tout.

La connaissance de la répartition des individus cibles parmi les diverses fréquences d'exposition est de première importance pour l'annonceur. Toutes choses égales par ailleurs, il est évident que plus le nombre d'individus cibles non exposés est faible, meilleur est le plan. Mais, la couverture nette n'est qu'un des aspects du problème car, parmi deux plans procurant la même couverture nette, l'un peut rassembler plus d'individus dans les hautes fréquences d'exposition que l'autre et, par conséquent, apporter davantage de répétitions. Dans le même ordre d'idée, si un annonceur connaît le niveau optimal de répétitions pour sa campagne, il essaiera de construire le plan d'insertion pour lequel le plus grand nombre d'individus cibles se trouvent au voisinage du niveau optimal de répétition. En effet, les individus exposés plus souvent le sont en pure perte tandis que ceux qui sont insuffisamment exposés ne sont pas convaincus. Le terme de couverture nette efficace a été mis en avant par l'agence J.W. Thompson (49) pour décrire le nombre d'individus cibles ayant reçu, au moins, trois expositions et, au plus, dix expositions. Bien que la fixation

des bornes d'efficacité entre 3 et 10 expositions soit discutable, puisque le niveau optimal dépend de la nature du produit, du message, de la durée et des objectifs de la campagne, elle a du moins le mérite d'expliciter la relation liant la distribution des occasions de voir aux objectifs de communication de la campagne. Le spécialiste de la recherche commerciale pourra même être en mesure de pondérer séparément chaque classe de fréquence en fonction de son expérience dans la vente du produit ou, mieux encore, sur la base d'expérimentations systématiques. Nous ne discuterons pas ici les mérites réciproques des pondérations et des courbes de réponses fondées sur le souvenir, les changements d'attitude ou les ventes, mais, quel que soit le critère choisi, la distribution des occasions de voir constitue le coeur de l'évaluation d'un plan publicitaire.

La distribution des occasions de voir s'obtient aisément à partir de la matrice d'exposition. Il suffit d'additionner les éléments de la matrice selon les diagonales montantes ainsi que l'illustre le diagramme suivant.

Dérivation de la Distribution des Occasions de Voir

à partir de la Matrice d'Exposition

(2 parutions de Time + 2 parutions de Newsweek)

Matrice d'Exposition

Exposition à Newsweek \ Exposition à Time	Zéro	Un	Deux
Zéro	95 355	9 591	8 525
Un	5 978	3 728	2 421
Deux	3 809	1 772	2 340

Distribution des Occasions de Voir

Individus	%	Fréquence
95 355	71,4	0 sur 4
15 569	11,7	1 sur 4
16 062	12,0	2 sur 4
4 193	3,1	3 sur 4
2 340	1,8	4 sur 4

La distribution des occasions de voir est le plus souvent présentée sous forme de pourcentage de la population totale. Elle permet de répondre aux trois questions suivantes : Combien d'individus ont été exposés K fois, au moins ? Combien d'individus ont été exposés K fois ou plus ? Combien d'expositions un individu moyen a-t-il reçues ? La réponse à la première question s'obtient en

Matrice de transition du Modèle Bêta Binomial

		N-1 ième Parution		Exposés à la Nième parution
		Non exposés	Exposés	
N ième parution	Non Exposés	$\frac{B\ (B+1)}{(A+B)\ (A+B+1)}$	$\frac{AB}{(A+B)\ (A+B+1)}$	$\frac{B}{A+B}$
	Exposés	$\frac{AB}{(A+B)\ (A+B+1)}$	$\frac{A\ (A+1)}{(A+B)\ (A+B+1)}$	$\frac{A}{A+B}$
Exposés à la N-1 ième parution		$\frac{B}{A+B}$	$\frac{A}{A+B}$	1,0

A partir de cette matrice de transition, nous pouvons calculer le coefficient de corrélation entre les audiences de deux parutions.

$$(50) \qquad \emptyset = \frac{1}{A+B+1}$$

Puisque ce coefficient dépend seulement des paramètres A et B, il reste donc constant quelles que soient les deux parutions considérées (i.e : N n'apparait pas dans l'équation (50)) et il ne se produit aucun apprentissage. Notons que, puisque les paramètres A et B sont par définition positifs, une corrélation négative entre deux parutions est impossible. Plus la somme des deux paramètres s'accroit, plus la corrélation entre les parutions diminue et plus le taux d'accumulation d'audience s'élève. Ce résultat s'accorde parfaitement avec l'interprétation précédente de la signification des paramètres. Une attraction élevée combinée à une répulsion élevée entraîne une modification constante des audiences et, par conséquent, un taux d'accumulation élevé. Inversement, si l'attraction comme la répulsion sont faibles, l'accumulation sera faible. A la limite, lorsque la somme des deux paramètres tend vers zéro, la corrélation entre les audiences de deux parutions tend à devenir parfaite et l'accumulation d'audience tend à devenir nulle. Cette situation caractériserait un support à audience parfaitement stable, n'attirant ni ne perdant un seul auditeur.

Il est intéressant de noter que le taux maximum d'accumulation permis par le modèle Bêta Binomial se produit lorsque la corrélation entre deux parutions devient nulle. Dans ce cas limite la distribution Bêta Binomiale se réduit à la distribution binomiale. En conséquence, le modèle Bêta Binomial n'est pas applicable quand le taux d'accumulation observé est supérieur à l'accumulation au pur hasard. Cette condition d'applicabilité peut être exprimée comme suit :

$$(51) \qquad R_1 < R_2 < 2R_1 - R_1^2$$

Si $R_2 = R_1$, l'audience est parfaitement stationnaire.

Si $R_2 = 2R_1 - R_1^2$, l'audience s'accumule au hasard et le modèle binomial s'applique.

Cependant, la condition d'applicabilité exprimée par la double inégalité (51) n'est guère restrictive et nous n'avons pas trouvé de support qui ne la vérifie pas.

Nous sommes maintenant prêts à discuter les diverses techniques permettant d'estimer les paramètres A et B.

a- Quand les seules données disponibles sont la couverture nette des deux premiers numéros, toutes les méthodes d'estimation deviennent identiques puisqu'il n'y a pas de degrés de liberté. Utilisant l'équation (39) nous obtenons :

$$\begin{cases} R_1 = E_1^1 = \dfrac{A}{A + B} \\[2ex] R_2 = E_1^2 + E_2^2 = \dfrac{2AB + A(A+1)}{(A+B+1)\ (A+B)} \end{cases}$$

En résolvant ces deux équations pour A et B, il en résulte :

$$(52) \qquad \begin{cases} \hat{A} = \dfrac{E_1^1\ (E_1^1 - E_2^2)}{E_2^2 - (E_1^1)^2} = \dfrac{R_1\ (R_2 - R_1)}{2R_1 - R_2 - (R_1)^2} \\[4ex] \hat{B} = \dfrac{\hat{A}\ (1 - E_1^1)}{E_1^1} = \dfrac{\hat{A}\ (1 - R_1)}{R_1} \end{cases}$$

Dans ces deux équations, les données R_1 et R_2 ou bien E_2^2 et E_1^1 sont exprimées en pourcentage de la population. Si l'on désire utiliser les valeurs absolues les deux équations deviennent :

$$\begin{cases} \hat{A} = \dfrac{E_1^1 (E_1^1 - E_2^2)}{POP * E_2^2 - (E_1^1)^2} = \dfrac{R_1 (R_2 - R_1)}{POP * (2R_1 - R_2) - (R_1)^2} \\[2ex] \hat{B} = \dfrac{\hat{A} (POP - E_1^1)}{E_1^1} = \dfrac{\hat{A} (POP - R_1)}{R_1} \end{cases}$$

En utilisant les équations (52) nous avons calculé les paramètres A et B pour nos 10 magazines habituels. Puis à l'aide des équations (40) et (39) nous avons calculé la couverture nette et la distribution des occasions de voir estimées par le modèle Bêta Binomial. Les deux tables suivantes montrent que les estimations du modèle sont particulièrement proches des données observées.

Pour la couverture nette, l'erreur type de prédiction n'est que 0,0076 en moyenne pour les dix magazines. C'est mieux que pour le modèle logarithmique et c'est pratiquement aussi bien que pour le modèle d'Agostini. Parmi les autres modèles stochastiques précédemment analysés, seul notre modèle à trois segments obtient un résultat comparable. Cependant, on doit remarquer que le modèle Bêta Binomial sous estime la couverture nette dans 9 cas sur 10.

La supériorité du modèle Bêta Binomial émerge clairement quand on se réfère à l'estimation de la distribution des occasions de voir. La valeur moyenne du Chi Deux n'est que de 1,9. C'est le meilleur résultat obtenu parmi tous les modèles stochastiques étudiés jusqu'ici. Pour sept magazines sur dix, le test du chi deux indique que la distribution estimée est identique à la distribution observée au seuil de confiance de 95 pour 100.

Le Modèle d'Accumulation Bêta Binomial

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	39,1	38,8	21,8	21,7	30,0	29,9	17,7	17,5	29,6	29,8
4	44,0	43,3	25,0	24,6	34,8	34,4	20,2	19,9	32,0	32,5
5	47,7	46,7	27,5	26,9	38,6	37,9	22,2	21,7	33,8	34,6
6	50,6	49,4	29,6	28,7	41,8	40,7	23,9	23,2	35,3	36,3
$\hat{A}$	0,329		0,161		0,331		0,119		0,146	
$\hat{B}$	1,160		1,201		1,859		1,119		0,614	
Erreur Type	0,0090		0,0058		0,0068		0,049		0,0070	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	49,4	49,1	32,6	32,2	41,2	40,8	32,2	32,1	52,5	52,2
4	54,1	53,4	36,5	35,6	46,0	45,3	35,6	35,4	56,6	55,9
5	57,7	56,6	39,6	38,2	49,7	48,7	38,3	38,0	59,7	58,7
6	60,5	59,1	42,1	40,3	52,7	51,4	40,4	40,0	62,2	60,8
$\hat{A}$	0,353		0,203		0,335		0,198		0,301	
$\hat{B}$	0,779		0,884		1,072		0,823		0,536	
Erreur Type	0,0094		0,0122		0,0089		0,0028		0,0093	

Le Modèle d'Accumulation Bêta Binomial

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	50,6	70,4	71,3	58,2	59,3	76,1	76,8	64,7	63,7
1	17,4	16,2	12,6	11,1	19,2	17,1	10,2	0,0	9,0	10,0
2	12,0	10,5	6,8	6,2	9,8	9,7	4,5	4,9	4,5	6,2
3	8,0	7,8	3,0	4,2	5,0	6,2	4,0	3,4	6,0	4,9
4	3,0	6,2	1,5	3,1	3,0	4,0	1,5	2,5	6,0	4,4
5	4,2	4,9	1,5	2,4	2,2	2,4	1,8	1,9	3,6	4,5
6	6,0	3,8	4,2	1,7	2,6	1,3	1,9	1,5	6,2	6,3
Chi Deux	4,6		4,5		1,7		1,1		1,6	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	40,9	57,9	59,7	47,3	48,6	59,6	60,0	37,8	39,2
1	16,7	15,0	15,1	12,5	17,8	16,1	12,4	12,2	15,0	12,8
2	11,0	10,6	8,2	7,7	11,3	10,5	8,6	7,6	9,5	9,2
3	9,4	8,8	4,9	5,9	7,3	8,1	4,9	5,8	6,9	7,9
4	6,7	8,0	5,2	5,0	6,4	6,6	5,2	5,0	8,5	7,8
5	6,6	7,8	2,5	4,5	4,4	5,5	4,0	4,6	6,8	8,7
6	10,1	8,9	6,2	4,7	5,5	4,6	5,3	4,8	15,5	14,4
Chi Deux	0,9		2,8		0,8		0,4		1,2	

b- Quand les données sont disponibles pour plus de deux parutions, on a le choix entre plusieurs méthodes d'estimation dont nous allons étudier les avantages et les inconvénients respectifs. En fait, il semble que seulement trois méthodes aient été utilisées jusqu'à présent, la méthode des moments, la méthode de la moyenne et du zéro, et la méthode des moindres carrés contraints. Nous présenterons ici deux autres méthodes d'estimation, la méthode de minimisation du chi deux et la méthode du maximum de vraisemblance. Toutes ces méthodes utilisent la distribution observée des occasions de voir comme point de départ.

1- La méthode des moments : Elle consiste à égaliser les moments de la distribution observée avec les moments de la distribution théorique et à résoudre les équations ainsi obtenues pour les paramètres inconnus. Puisque deux paramètres seulement doivent être estimés, il suffit de calculer la moyenne et la variance de la distribution observée des occasions de voir :

Moyenne $= \overline{K}_N = \Sigma_{K=0}^{N} K * O_K^N$

Variance $= \sigma_N^2 = \Sigma_{K=0}^{N} K^2 O_K^N - \overline{K}_N^2$

N = Nombre de parutions

O_K^N = Proportion d'individus exposés à K parutions parmi N.

La moyenne et la variance de la distribution Bêta Binomiale sont données par les formules (41) et (42). En égalisant la moyenne et la variance observée avec leur valeur théorique, nous obtenons :

$$\begin{cases} \dfrac{NA}{A+B} = \overline{K}_N \\[2ex] \dfrac{NAB\ (N+A+B)}{(A+B+1)\ (A+B)^2} = \sigma_N^2 \end{cases}$$

En résolvant ces deux équations pour les paramètres A et B, nous obtenons les estimateurs de la méthode des moments :

$$(53) \quad \begin{cases} \hat{A} = \overline{K}_N\ (Z-N)\ /\ \big(N(1-Z)\big) \\[2ex] \qquad\qquad \text{avec } Z = N\sigma_N^2\ /\ \big(\overline{K}_N\ (N-\overline{K}_N)\big) \\[2ex] \hat{B} = \hat{A}\ (N-\overline{K}_N)\ /\ \overline{K}_N \end{cases}$$

Ces estimateurs sont consistants et non biaisés, mais ils ne sont pas les plus efficaces (36). Ce qui signifie qu'il est possible de définir d'autres estimateurs dont l'erreur type d'estimation est plus faible. Cependant, les estimateurs de la méthode des moments sont faciles à calculer puisqu'ils n'impliquent pas le recours à une technique de minimisation. De plus, l'audience moyenne estimée d'une parution sera nécessairement égale à l'audience moyenne observée, ce qui n'est pas toujours le cas avec d'autres méthodes d'estimation.

2- La méthode de la moyenne et du zéro : Elle consiste à minimiser la différence entre la proportion observée d'individus non exposés et la proportion théorique, sous la contrainte d'égalisation de la moyenne théorique à la moyenne observée. Les estimateurs ainsi obtenus sont consistants et non biaisés. Cette méthode est pronée par Headen (37) sur la base des travaux d'Anscombe (38) qui a montré que la méthode est supérieure à la méthode des moments lorsque la distribution a la forme d'un J renversé avec une forte asymétrie positive. Cet argument est valide pour les magazines et la télévision, mais ne s'applique pas aux journaux quotidiens pour lesquels la distribution des occasions de voir présente une asymétrie négative. La formulation mathématique de la méthode est la suivante. Trouver les valeurs A et B qui minimisent :

$$(54) \quad \left[0_0^N - \frac{(B+N-1) \ldots (B)}{(A+B+N-1) \ldots (A+B)} \right]$$

sous la contrainte : $\bar{K}_N = NA / (A+B)$

La contrainte peut être introduite dans l'expression (54) en remplaçant le paramètre A par : $\bar{K}_N B / (\bar{N}-K_N)$. Puisque l'expression (54) n'est pas linéaire, nous avons utilisé l'algorithme de minimisation ZPOWEL fourni par le Software IMSL (39).

3- La méthode des moindres carrés : Elle consiste à trouver les valeurs A et B qui minimisent la somme du carré des différences entre les fréquences observées et les fréquences théoriques. On cherche donc les valeurs A et B qui minimisent :

$$(55) \quad \Sigma_{K=0}^N (0_K^N - E_K^N)^2 \qquad \text{avec } E_K^N \text{ donné par l'équation (39)}$$

Puisque l'expression (55) n'est pas linéaire, il faut recourir à un algorithme de minimisation tel que ZPOWEL, ou bien linéariser (55), par application du développement limité en série de Taylor, comme le conseillent Greene et

Stock (2). Nous avons choisi la première solution qui est la plus commode du point de vue des calculs. Les estimations obtenues sont consistantes et asymptotiquement non biaisées. Toutefois, l'audience moyenne estimée ne coïncide pas nécessairement avec l'audience moyenne observée. Pour remédier à cela, il est possible de recourir aux moindres carrés sous contrainte, en utilisant la technique du multiplicateur de Lagrange. Cependant, on observe que les estimations ainsi obtenues coïncident avec celles de la méthode des moments. En conséquence, nous présenterons seulement les estimations obtenues par la méthode des moindres carrés sans contrainte.

4- La méthode de minimisation du Chi Deux : Elle consiste à trouver les valeurs A et B qui minimisent la somme du carré des différences entre fréquences observées et fréquences théoriques, divisées par les fréquences théoriques. On cherche donc les valeurs A et B qui minimisent :

$$(56) \qquad \Sigma_{K=0}^{N} (O_K^N - E_K^N)^2 / E_K^N \qquad \text{avec } E_K^N \text{ défini par l'équation (39)}$$

Nous avons utilisé, ici aussi, l'algorithme ZPOWEL pour minimiser l'expression (56). Selon Fisher, l'estimateur du Chi Deux minimum est consistant et asymptotiquement non biaisé, bien qu'il ne soit pas l'estimateur le plus efficace. Cette propriété d'efficacité maximum n'est vérifiée que pour l'estimateur du maximum de vraisemblance.

5- La méthode du maximum de vraisemblance : On détermine tout d'abord la fonction de vraisemblance qui indique, pour chaque échantillon observé, sa probabilité. La fonction de vraisemblance de la distribution Bêta Binomiale n'est pas étudiée dans les ouvrages de statistiques mathématiques, nous allons donc procéder à sa dérivation.

Soit $L\ \{n_0, n_1, \ldots, n_K, \ldots, n_N\}$ la probabilité d'obtenir un échantillon de n individus tel que :

n_0 individus n'ont été exposés à aucune des N parutions,

n_1 individus ont été exposés à exactement une parution,

.

.

.

n_N individus ont été exposés à exactement N parutions parmi N, sachant que la probabilité d'obtenir exactement K expositions est donnée par la loi Bêta Binomiale. P(K/A,B,N) de l'expression (39).

Puisque tous les individus sont par hypothèse indépendants, la probabilité d'obtenir n_K individus exposés K fois sur N est donné par :

$$P(n_K \,/\, n, P(K/A,B,N)) = \left(P(K/A,B,N) \right)^{n_K}$$

puisqu'il existe $\dfrac{n\,!}{n_0\,!\,n_1\,!\,\ldots\,n_N!}$ façons d'affecter les n individus de l'échantillon, la fonction de vraisemblance est donc donné par la distribution multinomiale suivante :

$$L\,\{n_0, n_1, \ldots, n_N\} = \frac{n\,!}{n_0\,!\,n_1!\,\ldots\,n_N\,!} \prod_{K=0}^{N} \left(P(K/A,B,N)\right)^{n_K}$$

En prenant le logarithme de cette expression, la fonction de vraisemblance devient :

$$(57) \qquad \mathrm{Log}\,(L) = \mathrm{Cst.} + \Sigma_{K=0}^{N}\, n_K\, \mathrm{Log}\,\left(P(K/A,B,N)\right)$$

Ainsi l'estimation du maximum de vraisemblance s'obtient en cherchant les valeurs A et B qui maximisent l'expression (57). On cherche, donc, les valeurs de A et B qui maximisent la somme du produit du nombre d'individus exposés par le logarithme de la probabilité théorique d'exposition. Ces estimateurs sont consistants, asymptotiquement non biaisés et d'efficacité maximum.

La table suivante compare les valeurs de A et B obtenues par les cinq méthodes d'estimation, pour la distribution des occasions de voir du magazine Life en 1958. Cette table montre que les cinq méthodes conduisent à des estimations très proches des paramètres A et B.

Comparaison de 5 méthodes d'estimation

des paramètres du modèle Bêta Binomial

(Magazine Life - 1958 - Adultes de 20 ans ou plus)

Propriétés des Estimations

Méthode d'Estimation	A	B	Consistant	Non biaisé	Assympt. non biaisé	Efficacité
Moyenne et zéro	0,2599	0,9295	Oui	Oui	Non	Moyenne
Moments	0,2743	0,9810	Oui	Oui	Non	faible
Chi Deux Minimum	0,2793	0,9930	Oui	Non	Oui	forte
Maximum de vraisemblance	0,2789	0,9948	Oui	Non	Oui	Maximum
Moindres carrés	0,2827	1,0124	Oui	Non	Oui	Moyenne

La divergence maximum sépare la méthode de la moyenne et du zéro de la méthode des moindres carrés, mais elle n'est guère que de 9%. Cette petite différence au niveau des paramètres conduit, au niveau de la couverture nette, à une différence encore plus faible. Pour avoir une description plus précise du peu de divergence entre les diverses méthodes d'estimation, nous avons calculé pour chacune d'elles la couverture nette correspondant aux paramètres estimés. Les résultats sont présentés dans le tableau suivant.

Couverture Nette Estimée de 6 Parutions de Life

Méthode d'Estimation	Couverture Nette
Moyenne et zéro	46,0 %
Moments	46,7 %
Chi Deux minimum	47,0 %
Maximum de vraisemblance	47,0 %
Moindres carrés	47,1 %

La méthode des moindres carrés conduit à la plus forte couverture nette avec 47,1%, tandis que la méthode de la moyenne et du zéro conduit à l'estimation la plus faible avec 46,0%. La divergence maximum atteint donc à peine 1%. On peut donc conclure qu'en pratique la méthode d'estimation utilisée importe peu. Si l'absence de biais et la facilité des calculs sont considérés comme les aspects les plus importants, on peut se contenter de la méthode des moments. Si, par contre, l'efficacité maximum est requise, comme c'est le cas lorsque l'on cherche à tester une hypothèse, on préfèrera la méthode du maximum de vraisemblance. Quoiqu'il en soit, toutes les méthodes présentées sont invariantes au nombre de parutions considérées. Ce qui signifie que, si le processus d'exposition étudié est réellement Bêta Binomial, alors les paramètres estimés ne varient pas avec le nombre de parutions considérées.

Nous avons remarqué précédemment que le modèle Bêta Binomial, estimé sur la base des deux premières parutions, tend à sous estimer l'accumulation d'audience. Ce phénomène peut être maintenant mesuré avec précision en faisant varier le nombre de parutions prises en considération pour estimer les paramètres A et B. Le tableau suivant résume les résultats que nous avons obtenus pour le magazine Life.

Variation des Paramètres Estimés du modèle Bêta Binomial

en fonction du nombre de parutions

Nombre de Parutions	Méthode des Moments		Chi Deux Minimum		Méthode des Moindres carrés		Maximum de Vraisemblance	
	A	B	A	B	A	B	A	B
2	0,2743	0,9810	0,2743	0,9810	0,2743	0,9810	0,2743	0,9810
3	0,2743	0,9810	0,2749	0,9821	0,2756	0,9858	0,2748	0,9821
4	0,2743	0,9810	0,2764	0,9865	0,2787	0,9963	0,2762	0,9861
5	0,2743	0,9810	0,2780	0,9906	0,2810	1,0051	0,2776	0,9904
6	0,2743	0,9810	0,2793	0,9930	0,2827	1,0124	0,2789	0,9948

On observe que les estimations obtenues par la méthode des moments sont insensibles à la légère divergence qui existe entre le modèle Bêta Binomial et les données. C'est là un défaut assez grave de la méthode des moments. Au contraire, les trois autres méthodes d'estimation font apparaître la divergence entre le modèle et les données. Pour ces trois méthodes, les valeurs estimées des paramètres augmentent légèrement avec le nombre de parutions, ce qui est normal puisque nous avons déjà montré que les paramètres estimés sur la base de deux parutions seulement conduisent à une sous-estimation de la couverture nette.

Les cinq tables suivantes présentent, en détail, la distribution des occasions de voir en fonction du nombre de parutions, pour chacune des méthodes d'estimation. La dernière table démontre que le phénomène de variation des paramètres en fonction du nombre de parutions ne se limite pas au magazine Life mesuré par Politz en 1958, mais se vérifie aussi pour de nombreux autres magazines.

Distribution des Occasions de Voir pour 6 parutions selon 4 Méthodes d'Estimation

STEP 1 FIRST MOMENT ESTIMATES A = .2743 B = .9819 Life 1958, all adults

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 6	54000	52.97	54385	53.29	-.38
1 OUT OF 6	15600	15.29	14908	14.67	.62
2 OUT OF 6	9750	9.55	9972	9.38	.17
3 OUT OF 6	7000	6.86	7291	7.14	-.29
4 OUT OF 6	5250	5.14	6008	5.89	-.74
5 OUT OF 6	6000	5.88	5184	5.08	.80
6 OUT OF 6	4450	4.36	4645	4.55	-.19

MINIMUM CHI SQUARE ESTIMATES OF A AND B FOR LIFE 1958 ALL US ADULTS (20 AND OLDER)
USING ZXPOWEL ALGORITHM

A	B	MINCHI	ITERATION	ERROR
.27935	.99304	.26459	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 6	54000	52.92	54037	52.95	-.03
1 OUT OF 6	15600	15.29	15111	14.81	.48
2 OUT OF 6	9750	9.55	9680	9.49	.07
3 OUT OF 6	7000	6.86	7367	7.22	-.36
4 OUT OF 6	5250	5.14	6054	5.93	-.79
5 OUT OF 6	6000	5.88	5189	5.10	.78
6 OUT OF 6	4450	4.36	4607	4.51	-.15

LEAST SQUARE ESTIMATES FOR A AND B FOR LIFE 1958 ALL US ADULTS (20 AND OLDER)
USING ZXPOWEL ALGORITHM

A	B	LEAST SQ	ITERATION	ERROR
.28275	1.01238	.00016	4	0

SSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 6	54000	52.92	54003	52.92	-.00
1 OUT OF 6	15600	15.29	15238	14.93	.35
2 OUT OF 6	9750	9.55	9749	9.55	.00
3 OUT OF 6	7000	6.86	7395	7.25	-.39
4 OUT OF 6	5250	5.14	6044	5.92	-.78
5 OUT OF 6	6000	5.88	5145	5.04	.84
6 OUT OF 6	4450	4.36	4475	4.39	-.02

MAXIMUM LIKELIHOOD ESTIMATES FOR A AND B FOR LIFE 1958 ALL US ADULTS (20 AND OLDER)
USING ZXPOWEL ALGORITHM

A	B	ITERATION	ERROR
.27894	.99481	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 6	54000	52.92	54111	53.02	-.11
1 OUT OF 6	15600	15.29	15107	14.80	.48
2 OUT OF 6	9750	9.55	9430	9.49	.08
3 OUT OF 6	7000	6.86	7364	7.21	-.35
4 OUT OF 6	5250	5.14	6040	5.92	-.77
5 OUT OF 6	6000	5.88	5142	5.09	.80
6 OUT OF 6	4450	4.36	4543	4.49	-.13

Estimation des Paramètres par la Méthode des Moments

pour trois, quatre et cinq parutions

STEP 1 FIRST MOMENT ESTIMATES A = .2743 B = .9810
3 ISSUES

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 3	64100	62.81	64147	62.86	-.05
1 OUT OF 3	17850	17.49	17769	17.39	.14
2 OUT OF 3	11250	11.02	11391	11.16	-.14
3 OUT OF 3	8850	8.67	8803	8.63	.05

STEP 1 FIRST MOMENT ESTIMATES A = .2743 B = .9810
4 ISSUES

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 4	59850	58.65	60012	58.81	-.16
1 OUT OF 4	17000	16.66	16541	16.21	.45
2 OUT OF 4	10200	10.00	10606	10.39	-.40
3 OUT OF 4	8200	8.04	8118	7.95	.08
4 OUT OF 4	6800	6.66	6773	6.64	.03

STEP 1 FIRST MOMENT ESTIMATES A = .2743 B = .9810
5 ISSUES

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 5	56600	55.46	56879	55.74	-.27
1 OUT OF 5	16250	15.92	15662	15.35	.58
2 OUT OF 5	10000	9.80	10027	9.83	-.03
3 OUT OF 5	7000	6.86	7650	7.50	-.64
4 OUT OF 5	6750	6.61	6322	6.20	.42
5 OUT OF 5	5450	5.34	5509	5.40	-.06

MINIMUM CHI SQUARE ESTIMATES OF A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 3 ISSUES
USING ZXPOWEL ALGORITHM

A	B	MINCHI	ITERATION	ERROR
.27485	.98212	.00306	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF.	PERCENT
0 OUT OF 3	64100	62.81	64116	62.83		-.02
1 OUT OF 3	17850	17.49	17728	17.37		.12
2 OUT OF 3	11250	11.02	11402	11.17		-.15
3 OUT OF 3	8850	8.67	8804	8.63		.05

MINIMUM CHI SQUARE ESTIMATES OF A AND B FOR LILE 58 ALL US ADULTS 20 AND OLDER 4 ISSUES
USING ZXPOWEL ALGORITHM

A	B	MINCHI	ITERATION	ERROR
.27640	.98647	.02831	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF.	PERCENT
0 OUT OF 4	59850	58.65	59897	58.69		-.05
1 OUT OF 4	17000	16.66	16612	16.28		.38
2 OUT OF 4	10200	10.00	10650	10.44		-.44
3 OUT OF 4	8200	8.04	8136	7.97		.06
4 OUT OF 4	6800	6.66	6758	6.62		.04

MINIMUM CHI SQUARE ESTIMATES OF A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 5 ISSUES
USING ZXPOWEL ALGORITHM

A	B	MINCHI	ITERATION	ERROR
.27796	.99059	.10326	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF.	PERCENT
0 OUT OF 5	56600	55.46	56656	55.52		-.06
1 OUT OF 5	16250	15.92	15778	15.46		.46
2 OUT OF 5	10000	9.80	10106	9.90		-.10
3 OUT OF 5	7000	6.86	7698	7.54		-.68
4 OUT OF 5	6750	6.61	6338	6.21		.40
5 OUT OF 5	5450	5.34	5474	5.36		-.02

Estimation des Paramètres par la Méthode des Moindres Carrés

LEAST SQUARE ESTIMATES FOR A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 3 ISSUES
USING ZXPOWEL ALGORITHM

A	B	LEAST SQ	ITERATION	ERROR
.27577	.98582	.00000	3	136

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 3	64100	62.81	64039	62.81	.00
1 OUT OF 3	17850	17.49	17761	17.40	.09
2 OUT OF 3	11250	11.02	11410	11.18	-.16
3 OUT OF 3	8850	8.67	8786	8.60	.07

LEAST SQUARE ESTIMATES FOR A AND B FOR LILE 58 ALL US ADULTS 20 AND OLDER 4 ISSUES
USING ZXPOWEL ALGORITHM

A	B	LEAST SQ	ITERATION	ERROR
.27866	.99630	.00042	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 4	59850	58.65	59850	58.65	-.00
1 OUT OF 4	17000	16.66	16693	16.36	.30
2 OUT OF 4	10200	10.00	10686	10.47	-.48
3 OUT OF 4	8200	8.04	8131	7.97	.07
4 OUT OF 4	6800	6.66	6690	6.56	.11

LEAST SQUARE ESTIMATES FOR A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 5 ISSUES
USING ZXPOWEL ALGORITHM

A	B	LEAST SQ	ITERATION	ERROR
.28097	1.00513	.00008	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 5	56600	55.46	56602	55.46	-.00
1 OUT OF 5	16250	15.92	15887	15.57	.36
2 OUT OF 5	10000	9.80	10162	9.96	-.16
3 OUT OF 5	7000	6.86	7713	7.56	-.70

MAXIMUM LIKELIHOOD ESTIMATES FOR A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 3 ISSUES
USING ZXPOWEL ALGORITHM

A	B	ITERATION	ERROR
.27484	.98210	3	[illegible]

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 3	64100	62.81	64117	62.83	-.02
1 OUT OF 3	17850	17.49	17728	17.37	.12
2 OUT OF 3	11250	11.02	11402	11.17	-.15
3 OUT OF 3	8850	8.67	8803	8.63	.05

MAXIMUM LIKELIHOOD ESTIMATES FOR A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 4 ISSUES
USING ZXPOWEL ALGORITHM

A	B	ITERATION	ERROR
.27624	.98649	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 4	59850	58.65	59906	58.70	-.06
1 OUT OF 4	17000	16.66	16606	16.27	.39
2 OUT OF 4	10200	10.00	10646	10.43	-.44
3 OUT OF 4	8200	8.04	8134	7.97	.06
4 OUT OF 4	6800	6.66	6757	6.62	.04

MAXIMUM LIKELIHOOD ESTIMATES FOR A AND B FOR LIFE 58 ALL US ADULTS 20 AND OLDER 5 ISSUES
USING ZXPOWEL ALGORITHM

A	B	ITERATION	ERROR
.27764	.99038	4	0

ISSUES READ	OBSERVED	PERCENT	ESTIMATED	PERCENT	DIF. PERCENT
0 OUT OF 5	56600	55.46	56699	55.55	-.09
1 OUT OF 5	16250	15.92	15769	15.45	.47
2 OUT OF 5	10000	9.80	10098	9.90	-.10
3 OUT OF 5	7000	6.86	7691	7.54	-.68
4 OUT OF 5	6750	6.61	6333	6.21	.41
5 OUT OF 5	5450	5.34	5470	5.36	-.02

Estimation des Paramètres en fonction du Nombre de Parutions

Pour Divers Magazines

Comparaison des paramètres pour deux et six parutions

Magazine	Source	Année	Paramètres 2 parutions		Bêta 6 parutions		Différences	
			A	B	A	B	A	B
Life	Politz	1953	0,329	1,160	0,362	1,289	+0,033	+0,129
Life	Politz	1958	0,354	0,782	0,386	0,857	+o,032	+0,075
Look	Politz	1958	0,335	1,072	0,367	1,186	+0,032	+0,114
Look	Politz	1953	0,331	1,859	0,364	2,067	+0,033	+0,208
Ev. Post	Politz	1953	1,161	1,201	0,176	1,334	+0,015	+0,133
Ev. Post	Politz	1958	0,203	0,844	0,234	0,989	+0,034	+0,145
Lad.Ho.J.	Politz	1953	0,119	1,119	0,130	1,241	+0,011	+0,122
Lad.Ho.J.	Politz	1958	0,198	0,823	0,205	0,854	+0,007	+0,031
Rd. Digest	Politz	1958	0,301	0,546	0,328	0,589	+0,027	+0,043
This Week	Politz	1953	0,146	0,614	0,134	0,557	-0,012	-0,057

Comparaison des paramètres pour deux et quatre parutions

Magazine	Source	Année	Paramètres 2 parutions		Bêta 4 parutions		Différences	
			A	B	A	B	A	B
Life	Simmons	1964	0,188	0,364	0,166	0,317	-0,022	-0,047
Lad.Ho.J.	Simmons	1964	0,062	0,249	0,055	0,219	-0,007	-0,030
Rd. Digest	Simmons	1964	0,118	0,169	0,118	0,169	0,000	0,000
T.V. Guide	Simmons	1964	0,038	0,126	0,042	0,142	-0,004	+0,016
Fa. Circle	Simmons	1964	0,074	0,327	0,071	0,314	-0,003	-0,013
Gd. Hskp	Simmons	1964	0,078	0,330	0,073	0,306	-0,005	-0,024
Mc. Calls	Simmons	1964	0,085	0,291	0,066	0,223	-0,019	-0,068
Wom. Day	Simmons	1964	0,052	0,252	0,047	0,227	-0,005	-0,025

Les données d'audience que nous avons utilisé jusqu'ici ne concernaient que les magazines. Il ne faudait pas en déduire que le modèle Bêta Binomial ne s'applique qu'à cet unique media. Bien au contraire, chaque fois que nous avons pu disposer de données d'accumulation d'audience pour les journaux quotidiens, la télévision, la radio ou l'affichage, nous avons observé que le modèle Bêta Binomial s'ajustait fort bien aux distributions des occasions de voir, si différentes soient-elles. En fait, chaque media peut être décrit concisément et précisément par la répartition de ses audiences en terme de probabilité d'exposition, c'est-à-dire par sa fonction d'hétérogénéité f(p).

- Les audiences des media radiodiffusés sont concentrées autour des faibles probabilités d'exposition, avec un plus grand nombre d'individus dotés de probabilités moyennes ou fortes pour la télévision que pour la radio.

- A l'inverse, les audiences de l'affichage et des journaux quotidiens sont concentrées autour des fortes probabilités d'exposition, avec peu d'individus dotés de probabilités moyennes et très peu d'individus dotés de probabilités faibles.

- Les audiences des magazines occupent une position intermédiaire, caractérisée par une concentration des individus à chaque extrêmité de l'échelle des probabilités et peu d'individus au niveau des probabilités moyennes.

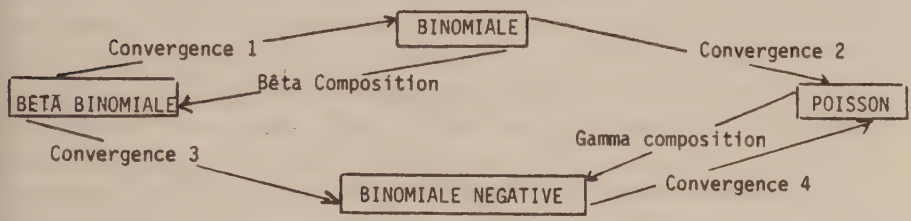
La table suivante illustre la forme type de la distribution des audiences qui caractérise chaque média, les positions relatives des paramètres A et B et les conséquences qui en résultent en matière de fidélité d'audience, de couverture nette cumulée et de fréquence moyenne d'exposition.

Description Synthétique de l'Audience des 5 Principaux Media

Media	Distribution des des Audiences	Valeurs Types des Paramètres	Fidélité	Couverture Nette	Fréquence Moyenne
RADIO		A ∠ 1 ∠ B	Faible	Faible	Faible
TELEVISION		A ∠ 1 ∠ B	Faible	Forte	Moyenne
MAGAZINES		A ∠ 1 ∠ B	Partagée	Moyenne	Moyenne
JOURNAUX		B ∠ A ∠ 1	Forte	Forte	Forte
AFFICHAGE		B ∠ A ∠ 1	Forte	Très forte	Très forte

Le Modèle d'accumulation Négatif Binomial.

La distribution Binomiale Négative (DBN) est une forme limite de la distribution Bêta Binomiale (DBB). Avant d'étudier son application à la modélisation de l'accumulation des audiences des supports radiodiffusés, nous présentons, dans le schéma suivant, les relations existant entre les distributions DBB, DBN, Binomiale et Poisson.



Nous avons déjà étudié le passage de la loi Binomiale à la loi Bêta Binomiale qui s'obtient en transformant la probabilité fixe p de la loi Binomiale en une variable aléatoire, Bêta distribuée, de paramètres A et B. Il est possible de faire la démarche inverse. On démontre (40) que si A et B tendent vers l'infini, tout en restant dans le rapport constant $p = A/(A + B)$, la loi Bêta Binomiale converge vers la loi Binomiale (convergence 1). On démontre aussi que la loi Binomiale converge vers la loi de Poisson lorsque le nombre de parutions tend vers l'infini tandis que p tend vers zéro, le produit np restant constant et égal à m (convergence 2). Greene et Stock (41) ont démontré que la loi Bêta Binomiale converge vers la loi Binomiale Négative quand le nombre de parutions tend vers l'infini tandis que la moyenne et la variance du nombre d'expositions restent constantes (convergence 3).

Le passage de la loi Bêta Binomiale à la loi Binomiale Négative conduit à remplacer l'exposition à une parution, qui est une variable aléatoire discrète, par l'exposition pendant un certain laps de temps, qui est une variable aléatoire continue. Par exemple, plutôt que de nous demander combien d'individus ont-ils exposés à K émissions télévisées parmi N, nous allons nous demander combien d'individus ont été exposés K fois pendant un laps de temps donné. Par conséquent, nous ne raisonnerons plus en termes de probabilité d'exposition, mais en termes de taux d'exposition par unité de temps.

Le modèle DBN considère que chaque individu possède un taux d'exposition personnel qui reste constant dans le temps et qui caractérise la

propension de cet individu à être exposé à un support publicitaire donné. Ce taux personnel d'exposition est indépendant du taux d'exposition des autres individus composant la population. Nous reconnaissons ici nos deux hypothèses familières de stationnarité dans le temps et de stationnarité inter individus. Il résulte, de ces hypothèses, que la distribution du nombre d'expositions auxquelles un individu sera soumis suit une loi de Poisson définie par le taux personnel d'exposition de cet individu. Bien entendu, les taux personnels d'exposition varient d'un individu à un autre. Cette hétérogénéité des individus est supposée suivre une distribution Gamma qui caractérise la population cible. On choisit la distribution Gamma plutôt que tout autre, car elle ne dépend que de deux paramètres a et b et car elle est la distribution conjuguée de la distribution de Poisson. En conséquence, le modèle d'accumulation Négatif Binomial n'est pas seulement un cas limite du modèle Bêta Binomial (convergence 3). Il est aussi une distribution composée résultant de la sommation de processus de Poisson dont le paramètre est Gamma distribué parmi les individus composant la population. A ce titre, on pourrait donc lui donner le nom de distribution Gamma Poisson.

Le modèle DBN a été utilisé avec succès par Ehrenberg (42) pour décrire les processus d'achats de biens de consommation courante. Son application au problème de l'accumulation des audiences a été suggérée récemment par Greene et Stock (41) et par Friedman (43). Ce dernier utilise le vocable de prédisposition à l'exposition en souvenir de la première application du modèle DBN par Greenwood et Yule(44) en 1920, dans le domaine de la prédisposition aux accidents.

Pour notre analyse du modèle DBN, nous adopterons la notation suivante :

u_i = Taux d'exposition personnel de l'individu i.

$f(u)$ = Distribution des taux d'expositions dans la population cible.

$P(K/u)$ = Probabilité d'obtenir exactement K expositions étant donné un taux d'exposition égal à u.

En vertu des hypothèses du modèle, $P(K/u)$ suit une distribution de Poisson de moyenne u :

$$P(K/u) = e^{-u} * \frac{u^K}{K!} \quad \text{avec : } K = 0, 1, \ldots, \infty \text{ ; entier}$$

L'hétérogénéité de la population, f(u), est représentée par une distribution Gamma de paramètres a et b :

$$f(u) = \frac{e^{-u/a} * u^{b-1}}{a^b * \Gamma(b)} \qquad \text{avec : } 0 \leq u < \infty \quad \text{et } a,\ b > 0 \text{ ; réels.}$$

En sommant les individus pour tous les taux d'expositions possibles, nous obtenons la distribution Gamma Poisson de paramètres a et b.

$$P\ (K/a,\ b) = \int_0^{\infty} P\ (K/u) * f(u)\ du$$

$$(58) \qquad P(K/a,b) = \frac{\Gamma(K + b)\ (1 + a)^{-b}}{\Gamma(K + 1)\ \Gamma(b)} * \left(\frac{a}{1 + a}\right)^K$$

$$\text{avec } K = 0, 1, \ldots, \infty \qquad \text{et } a, b > 0$$

Cette distribution donne la probabilité d'obtenir exactement K expositions, en fonction des paramètres a et b. Une expression rationnelle peut être dérivée de l'expression (58) en vue de faciliter les calculs :

$$(59) \qquad P(K/a,\ b) = (1 + a)^{-b}\ [a/\ (1 + a)]^K\ [(K + b - 1)\ \ldots(K)/\ K\ !]$$

Il convient de remarquer que dans les expressions (58) et (59), le nombre de parutions ou d'occasions de voir, N, a disparu. Nous pouvons le réintroduire en utilisant le concept d'impressions brutes.

$$(60) \qquad \left.\begin{array}{l}\text{Impressions brutes}\\ \text{après } N \text{ parutions}\end{array}\right\} = GI_N = m = \Sigma_{K=0}^{N} K * E_K^N = a * b$$

L'équation (60) rappelle que les impressions brutes sont par définition égales à la moyenne de la distribution des occasions de voir, m, qui est égale au produit des paramètres a et b dans le cas de la distribution binomiale négative. En conséquence, l'équation (58) peut être réécrite comme suit :

$$(61) \qquad P\ (K/m,b) = \frac{\Gamma(K + b)(1 + m/b)^{-b}}{\Gamma(K + 1)\ \Gamma(b)} * \left(\frac{m}{m + b}\right)^K$$

La couverture nette de N parutions est donnée par :

$$(62) \qquad R_N = 1 - P\ (0/m,\ b) = 1 - \left(\frac{b}{b + m}\right)^b$$

En se souvenant que $m = GI_N = N * R_1$, il vient :

$$R_N = 1 - \left(\frac{b}{b + NR_1} \right)^b$$

La variance de distribution Binomiale Négative est donnée par :

$$(63) \qquad \sigma^2 = m\,(a + 1) = ab\,(a + 1)$$

Finalement, tout comme la distribution DBB, la distribution DBN peut être calculée par récurrence :

$$(64) \qquad P\,(K/a, b) = \left(\frac{a}{a + 1} \right) \left(\frac{K + b - 1}{K} \right) P\,(K - 1 / a, b)$$

Les paramètres a et b de la distribution Binomiale Négative peuvent être estimés par l'une quelconque des cinq méthodes d'estimation présentées au cours de notre discussion du modèle Bêta Binomial. Ici encore, nous observons que toutes les méthodes donnent des estimations similaires des paramètres a et b. Nous nous contenterons donc de détailler seulement la méthode des moments et la méthode de la moyenne et du zéro qui sont les plus simples à mettre en oeuvre.

- <u>La méthode des moments</u>.

Soit m la moyenne et σ^2 la variance de la distribution observée des occasions de voir. En vertu des expressions (60) et (63), nous avons :

$$\begin{cases} m = ab \\ \sigma^2 = m\,(a + 1) \end{cases}$$

En résolvant ces deux équations pour les paramètres inconnus a et b, nous obtenons les estimateurs de la méthode des moments :

$$(65) \qquad \begin{cases} \hat{a} = (\sigma^2 - m) / m \\ \hat{b} = m^2 / (\sigma^2 - m) \end{cases}$$

<u>La méthode de la moyenne et du zéro</u>.

Anscombe (38) a démontré que l'estimateur du maximum de vraisemblance

du produit des paramètres ab est la moyenne m de la distribution observée des occasions de voir. Cet estimateur est non biaisé. Le paramètre "a" est estimé en égalisant la proposition observée d'individus non exposés, O_o^N , avec la proposition théorique donnée par l'équation (61) :

$$O_o^N = (1 + m/b)^{-b} = (1 + a)^{-m/a}$$

Cette équation peut être résolue par la méthode itérative suivante :

$$(66) \qquad \hat{a}_{i+1} = c \log (1 + \hat{a}_i) \qquad \text{avec } c = - m / \log (O_o^N)$$

Il faut noter que si c est inférieur à l'unité, il en résulte une valeur négative de a, ce qui est contraire aux hypothèses du modèle. Par conséquent, la condition d'applicabilité du modèle DNB est la suivante :

$$(67) \qquad m > - \log (O_o^N)$$

Lorsqu'une exactitude parfaite n'est pas requise, l'équation (66) peut être remplacée par la table fournie par Chatfield (45) qui donne la valeur de "a" avec trois décimales, en fonction des valeurs de c.

Lorsque les seules données disponibles sont l'audience des deux premières parutions, les paramètres a et b pourront être estimés comme suit :

Nous avons : $E_o^2 = 1 - R_2$ et $GI_2 = 2 R_1 = m$

Nous en déduisons la valeur de c : $c = - 2 R_1 / \text{Log} (1 - R_2)$

La condition d'applicabilité est vérifiée quand : $2 R_1 > - \text{Log} (1 - R_2)$

Dans ce cas, le paramètre "a" est donné par l'équation (66) et le paramètre b est égal à $\hat{b} = 2 R_1/\hat{a}$

La table suivante compare, pour nos dix magazines habituels, les estimations des paramètres a et b, obtenus par la méthode de la moyenne et du zéro, en fonction du nombre de parutions considérées. Nous comparons aussi, dans le cas de six parutions, les estimations de la méthode des moments avec celles de la méthode de la moyenne et du zéro. Le test d'ajustement du Chi Deux nous permet de déterminer quelle est la méthode qui reproduit le mieux la

distribution observée des occasions de voir.

Estimation des Paramètres de la Distribution Binomiale Négative

Magazines	Méthode de la Moyenne et du Zéro								Méthode des moments					
	2 Parutions		3 Parutions		4 Parutions		5 Parutions		6 Parutions			6 Parutions		
	a	b	a	b	a	b	a	b	a	b	X^2	a	b	X^2
LIFE 1953	0,26	1,70	0,74	0,90	1,20	0,73	1,60	0,66	2,17	0,61	5,0	1,42	0,94	8,
POST 1953	0,43	0,55	0,99	0,36	1,51	0,31	2,04	0,29	2,56	0,28	3,3	2,04	0,36	4,
LOOK 1953	0,21	1,45	0,58	0,78	0,92	0,65	1,27	0,60	1,60	0,57	1,4	1,32	0,70	2,
LHJ 1953	0,50	0,38	1,09	0,26	1,67	0,23	2,26	0,21	2,83	0,20	1,8	1,92	0,30	4,
WEEK 1953	0,61	0,63	1,51	0,38	2,49	0,31	3,50	0,27	4,52	0,25	10,1	2,10	0,55	28,
LIFE 1958	0,25	2,50	0,83	1,13	1,41	0,89	1,98	0,79	2,56	0,73	8,2	1,30	1,44	15,
POST 1958	0,45	0,87	1,08	0,54	1,69	0,46	2,29	0,42	2,90	0,40	4,7	1,78	0,65	9,
LOOK 1958	0,26	1,83	0,76	0,94	1,26	0,76	1,75	0,68	2,24	0,64	4,0	1,34	1,06	8,
LHJ 1958	0,46	0,85	1,14	0,51	1,84	0,42	2,54	0,38	3,26	0,36	5,5	1,81	0,65	13,
DIGEST 1958	0,31	2,32	1,02	1,06	1,74	0,83	2,46	0,73	3,17	0,68	14,5	1,38	1,56	28,

Le test du Chi Deux montre que la méthode de la moyenne et du zéro est toujours préférable à la méthode des moments. Nous remarquons aussi que la valeur des paramètres a et b se modifie énormément selon le nombre de parutions considérées et ceci est vrai pour tous les magazines examinés. Cela signifie que le modèle DBN ne s'applique pas à l'accumulation d'audience des magazines.

Nous pouvons le démontrer plus clairement encore à l'aide des deux tables suivantes qui comparent la couverture nette et la distribution des occasions de voir estimées par le modèle DBN avec la couverture nette et la distribution observée.

Ces tables montrent que, même estimé sur la base de six parutions, le modèle DNB ne reproduit pas de façon satisfaisante la couverture nette cumulée. Le modèle sous-estime systématiquement la couverture nette. L'erreur type de

Le Modèle d'Accumulation Négatif Binomial
Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
1	22,1	17,2	11,8	9,4	15,1	12,5	9,6	7,6	19,2	13,3
2	32,4	28,3	17,8	15,7	23,9	21,5	14,4	12,6	25,9	20,9
3	39,1	36,2	21,8	20,4	30,0	28,3	17,7	16,4	29,6	26,0
4	44,0	42,1	25,0	24,1	34,8	33,7	20,2	19,4	32,0	29,8
5	47,7	46,8	27,5	27,1	38,6	38,1	22,2	21,8	33,8	32,8
6	50,6	50,6	29,6	29,6	41,8	41,8	23,9	23,9	35,3	35,3
$\hat{a}$	2,167		2,562		1,597		2,834		4,521	
$\hat{b}$	0,612		0,276		0,567		0,203		0,255	
Erreur type	0,0179		0,0088		0,0102		0,0078		0,0216	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Parutions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
1	31,2	22,9	19,4	14,6	23,8	18,3	19,4	14,4	36,0	25,1
2	42,6	36,3	27,4	23,8	34,4	29,9	27,3	23,1	46,5	38,8
3	49,4	45,3	32,6	30,2	41,2	38,0	32,2	29,2	52,5	47,6
4	54,1	51,7	36,5	35,1	46,0	44,1	35,6	33,8	56,6	53,9
5	57,7	56,3	39,6	38,9	49,7	48,9	38,3	37,4	59,7	58,5
6	60,5	60,5	42,1	42,1	52,7	52,7	40,4	40,4	62,2	62,2
$\hat{a}$	2,557		2,889		2,244		3,260		3,171	
$\hat{b}$	0,732		0,402		0,636		0,357		0,681	
Erreur type	0,0243		0,0143		0,0188		0,0181		0,0285	

Le Modèle d'Accumulation Négatif Binomial
Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		LADIES H.J.		THIS WEEK	
Expositions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
0	49,4	49,4	70,4	70,4	58,2	58,2	76,1	76,1	64,7	64,7
1	17,4	20,7	12,6	14,0	19,2	20,3	10,2	11,4	9,0	13,5
2	12,0	11,4	6,8	6,4	0,8	9,8	4,5	5,1	4,5	6,9
3	8,0	6,8	3,0	3,5	5,0	5,1	4,0	2,8	6,0	4,3
4	3,0	4,2	1,5	2,1	3,0	2,8	1,5	1,6	6,0	1,0
5	4,2	2,6	1,5	1,3	2,2	1,6	1,8	1,0	3,6	2,0
6										
Chi Deux	5,0		3,3		1,4		1,8		10,1	

Enquête 1958	LIFE		POST		LOOK		LADIES H.J.		DIGEST	
Expositions	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.	OBS.	EST.
0	39,5	39,5	57,9	57,9	47,3	47,3	59,6	59,6	37,8	37,8
1	16,7	20,8	15,1	17,3	18,8	20,8	12,4	16,3	15,0	19,6
2	11,0	12,9	8,2	9,0	11,3	11,8	8,6	8,5	9,5	12,5
3	9,4	8,5	4,9	5,4	7,3	7,2	4,9	5,1	6,9	8,5
4	6,7	5,7	5,2	3,4	6,4	4,5	5,2	3,3	8,5	6,0
5	6,6	3,9	2,5	2,2	4,4	2,9	4,0	2,2	6,8	4,2
6	10,1	2,7	6,2	1,5	5,5	1,9	5,3	1,5	15,5	3,0
Chi Deux	8,2		4,7		4,0		5,5		14,5	

prédiction est en moyenne de 0,0170 pour les dix magazines. Bien évidemment, les résultats sont encore pire lorsque les estimations ne sont basées que sur les deux premières parutions. Dans ce cas, le modèle surestime systématiquement la couverture nette et l'erreur type de prédiction atteint 0,0815. Ce résultat est l'un des plus mauvais parmi tous les modèles stochastiques à l'exception du modèle binomial et du modèle d'hétérogénéité linéaire.

Il est évident que le modèle DBN ne s'applique pas pour les magazines et la comparaison des occasions de voir explique pourquoi. Les fréquences binomiales négatives décroissent de façon monotone, de l'absence d'exposition jusqu'à six expositions. Ainsi, le modèle DBN ne prend pas en considération les lecteurs fidèles qui sont systématiquement sous-estimés. La raison de cet échec tient à la forme de la distribution Gamma. Cette distribution peut prendre une forme en cloche ou une forme décroissante, mais elle ne peut pas prendre une forme en U. En conséquence, le modèle DBN ne convient pas aux magazines, pour lesquels l'hétérogénéité de la population prend toujours une forme en U.

Bien qu'il échoue pour les magazines, le modèle DBN donne des résultats satisfaisants pour les supports publicitaires radiodiffusés qui sont caractérisés par un grand nombre d'auditeurs dans la zone des basses probabilités d'expositions et un petit nombre d'auditeurs fidèles. Néanmoins, même dans ce cas, le modèle DBN présente un inconvénient logique mineur par rapport au modèle Bêta Binomial. Du fait que la distribution Binomiale Négative est semi infinie (i.e 0 K ∞), la somme, de zéro jusqu'à N parutions, de la distribution des occasions de voir n'est pas égale à l'unité. D'ordinaire, la divergence est extrêmement faible et pratiquement négligeable puisque les probabilités telle que K - N sont proches de zéro. Néanmoins, il convenait de noter que, dans le modèle DBN, K varie de zéro à l'infini, alors que les occasions de voir réelles ne varient que de zéro à N.

Deux applications du modèle DBN à l'estimation de l'accumulation des audiences télévisées ont été menées à bien avec succès. Une par le département de recherche opérationnelle de J. Walter Thompson en Grande-Bretagne sous la direction de John Davis (47), et l'autre par l'auteur pour le compte de la Société Interactive Market Systems (48) aux Etats-Unis. Les données utilisées par l'équipe de JWT représentaient 300 programmes ITV fournies par la société de mesure d'audience JICTAR. Les données utilisées dans l'étude IMS proviennent de mesures d'audiences télévisées effectuées par Nielsen (48). 1202 hommes et 1338 femmes ont rempli des livres d'écoute pour une période de 12 semaines, du 27 septembre à la fin décembre 1971. Pour un total de 278 programmes, les données suivantes ont été tabulées ; nombre de programmes, nombre de parutions à l'écran, nombre de chaînes, audience cumulée pour quatre semaines, audiences

brutes et distribution des occasions de voir. De plus les programmes étaient classifiés, selon l'heure, en programmes de la journée, de la soirée, de début de soirée, de fin de soirée et en programmes sportifs. La table suivante donne une description détaillée de cette banque de données.

Pour tous ces 278 plans publicitaires télévisés, à l'exception d'un seul, la distribution Binomiale Négative reproduit de façon très satisfaisante la distribution observée des occasions de voir. En fait, le test du Chi Deux conduit à accepter, au seuil de confiance de 95 / 100, l'hypothèse que toutes les distributions observées, sauf une, sont des distributions Binomiales Négatives.

L'étude de J.W.T. ne présente aucun test statistique de la qualité des ajustements obtenus, mais une série de tables permettant de comparer les fréquences estimées aux fréquences observées. L'examen de ces tables fait apparaître là aussi, un excellent ajustement du modèle aux données. Deux plans publicitaires, tirés au hasard dans chacune des deux études sont reproduits ci-dessous :

L'Ajustement du modèle Binomial Négatif à deux plans publicitaires types

Nombre d'expositions	Etude J. W. T. Observé	Etude J. W. T. Estimé	Etude I. M. S. Observé	Etude I. M. S. Estimé
1	17,2	18,2	16,2	17,1
2	14,6	16,7	13,6	14,1
3	15,3	13,6	10,4	11,3
4	8,9	10,5	8,4	8,9
5	7,6	7,8	5,8	6,9
6	8,3	5,6	7,1	5,3
7	5,1	3,9	6,5	4,1
8	3,8	2,8	2,6	3,2
9 ou plus	3,9	4,1	10,2	7,9
Couverture nette	84,7	83,2	80,8	78,8
Audience brute	320		360	

Un aussi bon ajustement ne peut être le fruit du hasard. Il confirme les hypothèses sur lesquelles le modèle Binomial Négatif est construit.

Nous avons maintenant terminé notre présentation des modèles

Description de la banque de Données

94 Plans Publicitaires Télévisés Simples

classés selon l'heure d'antenne

Heure d'antenne	Nombre de plans	Nombre de parutions	Nombre de programmes	Etendue des audiences brutes
Soirée	32	1 à 24	1 à 18	18 à 434
Journée	28	6 à 24	2 à 12	32 à 311
Sports	16	4 à 10	1 à 6	54 à 144
Début de soirée	10	2 à 12	1 à 4	25 à 141
Fin de soirée	8	2 à 12	1 à 2	12 à 54

184 Plans Publicitaires Mixtes

classés selon les combinaisons d'heure d'antenne

Combinaisons d'heure d'Antenne	Nombre de Plans	Nombre de Parutions	Etendue des audiences brutes
1/3 journée + 2/3 soirée	28	8 à 64	66 à 605
1/2 journée + 1/2 soirée	18	7 à 60	50 à 528
2/3 journée + 1/3 soirée	18	13 à 52	107 à 384
1/3 début soirée + 2/3 soirée	22	5 à 28	72 à 435
1/3 sport + 2/3 soirée	20	10 à 24	164 à 415
2/3 sport + 1/3 soirée	16	5 à 20	74 à 328
Sport + début soirée + soirée	16	10 à 32	149 à 469
1/2 début soirée + sport	14	16	194 à 197
1/3 début soirée + 2/3 journée	10	15 à 30	127 à 229
Journée + début soirée + soir.	10	20 à 40	192 à 384
Début S. + soirée + fin de s.	8	10 à 38	146 à 478
Journée + soirée + fin soirée	6	18 à 60	175 à 522
Début de soirée + soirée + (fin de soirée + journée)	4	34 à 64	297 à 559
2/3 début s. + 1/3 fin soirée	4	12 à 24	98 à 194

d'accumulation fondé sur l'hypothèse de l'hétérogénéité des individus. Nous allons maintenant étudier une nouvelle catégorie de modèles qui combine l'hétérogénéité des individus et l'hétérogénéité des segments.

324. Les Modèles d'Accumulation à Double Hétérogénéité

Hypothèses

a- La population comprend plusieurs segments. Chaque segment est caractérisé par sa propre distribution des probabilités individuelles d'exposition.

b- Les probabilités d'exposition sont stationnaires d'une parution à une autre et d'un individu à un autre.

Le Modèle d'Accumulation de Broadbent

Le premier modèle d'accumulation à double hétérogénéité a été présenté en 1964 par Broadbent (27). Ce modèle suppose l'existence de quatre segments d'audience :

- Les non lecteurs dont la probabilité d'exposition est nulle et qui représentent un pourcentage W_0 de la population cible.
- Les lecteurs absolument fidèles dont la probabilité d'exposition est un et qui représentent un pourcentage W_1 de la population cible.
- Les lecteurs occasionnels qui représentent un pourcentage W_2 de la population cible et qui sont distribués selon une fonction qui décroîtavec la probabilité d'exposition.
- Les lecteurs réguliers qui représentent un pourcentage W_3 de la population cible et qui sont distribués selon une fonction qui croît avec la probabilité d'exposition.

Les deux premiers segments sont homogènes tandis que les deux derniers sont distribués selon une fonction type puissance de paramètre r. La fonction d'htérogénéité qui en résulte est la suivante :

$$(68) \qquad f(p) = \begin{cases} W_0 & \text{pour } p = 0 \\ W_2\,(1+r)(1-p)^r + W_3\,(1+r)\,p^r & \text{pour } 0 < p < 1 \\ W_1 & \text{pour } p = 1 \end{cases}$$

Le modèle de Broadbent dépend de quatre paramètres, W_0, W_1, W_2 et r, puisque W_3 est donné par : $W_3 = 1 - W_0 - W_1 - W_2$. En conséquence, l'ajustement du modèle requiert la connaissance de données d'audience portant sur quatre parutions au moins. En sommant les individus, pour toutes les probabilités d'exposition possibles à l'exception de p = 0 et p = 1, on obtient, sur la base de l'équation (32), la distribution des occasions de voir suivante :

$$E_K^N = \binom{N}{K}(1+r) \int_0^1 p^K(1-p)^{N-K}\left(W_2(1-p)^r + W_3\, p^r\right) dp$$

$$E_K^N = \left(W_2(1+r)\binom{K}{K} \int_0^1 p^K(1-p)^{N-K+r} dp\right) + \left(W_3(1+r)\binom{N}{K}\int_0^1 p^{K+r}(1-p)^{N-K} dp\right)$$

$$(69) \qquad E_K^N = (1+r)\binom{N}{K} * \frac{W_2\, \Gamma(K+1)\Gamma(N+r-K+1) + W_3\, \Gamma(r+K+1)\Gamma(N-K+1)}{\Gamma(N+r+1)} \qquad 0 < K < N$$

$$E_0^N = W_0$$

$$E_N^N = W_1$$

Sur la base de cette équation, il est possible d'exprimer le paramètre r en fonction de distribution des occasions de voir. Pour le cas de quatre parutions nous obtenons :

$$(70) \qquad \frac{3\,(r+2)}{r^2 + 8r + 18} = \frac{E_2^4}{E_1^4 + E_2^4 + E_3^4}$$

L'équation (70) peut être résolue pour r par une méthode itérative et sous la contrainte que r soit positif. Les trois paramètres restants, W_0, W_1, W_3, sont alors estimés en égalisant l'audience moyenne, la fréquence nulle, et la fréquence 4 expositions sur 4,avec leurs valeurs théoriques tirées de l'équation (69)

Ce modèle présente des difficultés du point de vue du calcul numérique et de fait, l'illustration qui en est donnée par Broadbent se limite au cas où le paramètre r est un nombre entier. L'introduction d'un segment de non lecteurs est intéressante puisqu'elle conduit à une couverture nette toujours inférieure à la population cible quel que soit le nombre de parutions. Par contre, l'introduction d'un segment de lecteurs absolument fidèles est fort

discutable. En effet, si un tel groupe existe, sa taille doit être négligeable dès qu'un nombre suffisant de parutions est considéré . En fait, dans l'illustration qu'il donne de son modèle, la proportion estimée de lecteurs absolument fidèles n'est que de 3%. Afin de réduire les difficultés de calcul liées à l'estimation des paramètres ainsi que le nombre de parutions exigées, nous avons ajusté le modèle de Broadbent, sur la base des deux premières parutions seulement, en supprimant le segment des lecteurs absolument fidèles et le segment des non lecteurs. Dans ce cas, l'équation (69) se réduit à :

$$(71) \qquad E_K^N = (1+r)\binom{K}{N}\left(W\beta(K+r+1,N-K+1) + (1-W)\beta(K+1,N+r-K+1)\right)$$

Le paramètre W, qui mesure la proportion de lecteurs réguliers, et le paramètre r, qui détermine la forme de f(p), peuvent être estimés à partir de la distribution des occasions de voir pour deux parutions. Avec N=2 et K =1, les deux Bêta coefficients sont égaux et (71) se réduit à :

$$(72) \qquad E_1^2 = \frac{2\ (r+1)}{(r+2)(r+3)} = 2\ (R_2 - R_1)$$

En résolvant l'équation (72) pour r et en égalisant l'audience observée d'une parution avec sa valeur théorique, nous obtenons :

$$(73) \qquad \hat{W} = \left(E_1^1(\hat{r}+2) - 1\right)/\hat{r}$$

Sur la base des équations (72) et(73), nous avons estimé r et W pour nos 10 magazines habituels. Ensuite, à l'aide de l'équation (71) nous avons calculé la couverture nette et la distribution des occasions de voir estimées par le modèle de Broadbent.

A l'examen de la première table, qui compare la couverture nette estimée avec la couverture nette réelle, il apparaît que le modèle de Broadbent à deux segments surestime systématiquement la couverture nette. L'erreur type moyenne est de 0,0295. Ce résultat est inférieur à ceux obtenus par le modèle à trois segments de l'auteur et par le modèle Bêta Binomial. Néanmoins, le modèle de Broadbent est supérieur au modèle d'accumulation à deux segments homogènes. Ce résultat était attendu puisque comme nous l'avons déjà mentionné, il n'y a aucune raison a priori pour que les individus appartenant à un même segment possèdent une probabilité identique d'exposition.

Le Modèle d'Accumulation à Deux Segments de Broadbent

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	39,1	39,8	21,8	22,8	30,0	30,8	17,7	18,6	29,6	31,0
4	44,0	45,7	25,0	27,2	34,8	34,8	20,2	22,3	32,0	35,4
5	47,7	50,5	27,5	31,2	38,6	41,4	22,2	25,7	33,8	39,3
6	50,6	54,5	29,6	34,7	41,8	45,5	23,9	28,9	35,3	42,7
$\hat{r}$	5,396		12,518		7,117		16,722		10,755	
$\hat{w}$	0,118		0,057		0,529		0,477		0,135	
Erreur type	0,0254		0,0336		0,0253		0,0325		0,0497	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	49,4	49,9	32,6	33,4	41,2	41,9	32,2	33,3	52,5	52,9
4	54,1	55,4	36,5	38,5	46,0	47,7	35,6	38,3	56,6	57,7
5	57,7	59,8	39,6	42,8	49,7	52,4	38,3	42,6	59,7	61,5
6	60,5	63,3	42,1	46,6	52,7	56,4	40,4	46,3	62,2	64,7
$\hat{r}$	4,402		8,285		5,106		8,446		5,201	
$\hat{w}$	0,227		0,120		0,135		0,122		0,306	
Erreur type	0,0189		0,0295		0,0247		0,0390		0,0166	

Le Modèle d'Accumulation à Deux Segments de Broadbent

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	45,5	70,4	65,3	58,2	54,5	76,1	71,1	64,7	57,3
1	17,4	24,0	12,6	21,2	19,2	24,9	10,2	18,8	9,0	20,5
2	12,0	11,8	6,8	6,1	9,8	10,3	4,5	4,3	4,5	6,6
3	8,0	5,6	3,0	1,6	5,0	3,9	4,0	0,9	6,0	2,0
4	3,0	3,3	1,5	0,6	3,0	1,7	1,5	0,3	6,0	1,4
5	4,2	3,7	1,5	1,3	2,2	1,6	1,8	1,0	3,6	3,3
6	6,0	6,1	4,2	3,9	2,6	3,1	1,9	3,6	6,2	8,9
Chi Deux	3,7		7,5		3,0		12,7		23,9	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	36,7	57,9	53,4	47,3	43,6	59,6	53,7	37,8	35,3
1	16,7	21,3	15,1	22,5	17,8	23,7	12,4	22,3	15,0	19,1
2	11,0	11,9	8,2	8,6	11,3	12,0	8,6	8,4	9,5	9,9
3	9,4	6,9	4,9	3,1	7,3	5,9	4,9	3,0	6,9	5,8
4	6,7	5,5	5,2	1,9	6,4	3,7	5,2	1,9	8,5	5,6
5	6,6	6,9	2,5	3,2	4,4	4,2	4,0	3,2	6,8	8,7
6	10,1	10,8	6,2	7,3	5,5	6,9	5,3	7,5	15,5	15,6
Chi Deux	2,5		7,1		4,0		12,4		3,0	

A l'examen de la seconde table, nous constatons que le modèle de Broadbent reproduit assez bien la distribution des occasions de voir. Les distributions estimées possèdent la forme en U désirée, mais il se produit une surestimation aux deux extrêmités de la distribution et une sous estimation au centre de la distribution. La valeur moyenne du test du Chi Deux est de 8, ce qui est un résultat moins satisfaisant que pour le modèle Bêta Binomial.

Le second modèle recourant à une double hétérogénéité de la population a été mis au point par Morgensztern (28) en 1970. Il fait l'objet de notre prochaine analyse.

Le Modèle d'Accumulation de Morgensztern

Morgensztern s'attache à déterminer, parmi les individus exposés à la N+1 ième parutions, le nombre D_N d'individus préalablement exposés à l'une des N parutions précédentes. Nous allons illustrer ce concept à l'aide de données chiffrées. Supposons que la couverture nette cumulée de 5 parutions d'un support soit la suivante :

Parutions	1	2	3	4	5
Couverture nette	250	345	410	460	500

Supposons de plus que l'audience de chaque parution demeure constante et égale à 250. La seconde parution apporte 95 lecteurs nouveaux (i.e. : 345-250) qui n'étaient pas exposés à la première parution. Donc parmi les 250 lecteurs de la deuxième parution, 155 lecteurs (i.e. : 250-95) ont été aussi exposés à la première parution et, par conséquent, D_1 = 155. La troisième parution apporte 65 lecteurs nouveaux qui ne furent exposés ni à la première ni à la seconde parution. Ainsi, les 185 lecteurs restant de la troisième parution ont été nécessairement exposés à l'une au moins des deux parutions précédentes. Il en résulte que D_2 = 185. D'une façon générale, si nous appelons A_N l'audience de la N ième parution et Δ_N l'accroissement de la couverture nette consécutif à la N+1 ième parution, nous avons :

$$D_N = A_N - \Delta_N = A_N - (R_{N+1} - R_N)$$

Si nous acceptons l'hypothèse de la constance de l'audience de chaque parution, c'est-à-dire $A_N = A_{N+1} = A_1 = R_1$, nous obtenons :

$$D_N = R_1 - (R_{N+1} - R_N) \tag{74}$$

Avec les données de notre exemple, les quatre premières valeurs de D_N sont égales à :

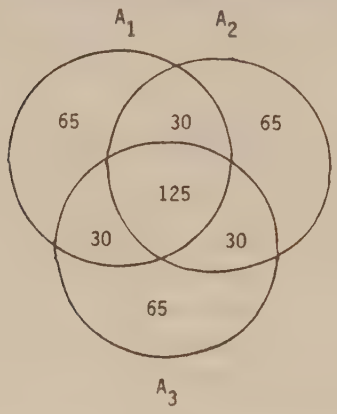
Parutions	1	2	3	4
D_N	155	185	200	210

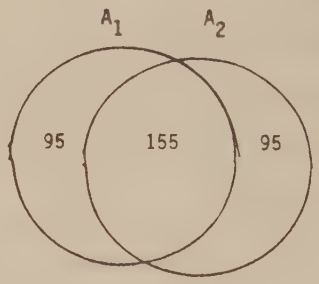
Morgensztern cherche ensuite à relier la couverture nette cumulée à la distribution des occasions de voir par l'intermédiaire de D_N.

Il propose la relation suivante :

$$(75) \quad E_K^N = 2R_1 - (D_K + D_{N-K}) = (R_{K+1} - R_K) + (R_{N-K+1} - R_{N-K})$$

Cependant, cette formule est inexacte, comme le prouve le simple exemple suivant qui considère trois parutions A_1, A_2 et A_3 ayant chacune une audience de 250 individus pour une population cible de 100 individus.





Couverture Nette	Distribution des Occasions de Voir		D_N
	3 parutions	2 Parutions	
$R_1 = 250$	$E_3^3 = 125$	$E_2^2 = 155$	$D_1 = 155$
$R_2 = 345$	$E_2^3 = 90$	$E_1^2 = 190$	$D_2 = 185$
$R_3 = 410$	$E_1^3 = 195$	$E_0^2 = 655$	$D_3 = 200$
	$E_0^3 = 590$		

Sur la base de la formule (75) de Morgensztern, nous devrions avoir :

- Pour deux parutions :

$E_2^2 = 2R_1 - D_2 - D_0 = 500 - 185 = 315 \neq 155$ Erreur

$E_1^2 = 2R_1 - D_1 - D_1 = 500 - 155 - 155 = 190 = 190$ Exact

$E_0^2 = 2R_1 - D_0 - D_2 = 500 - 185 = 315 \neq 655$ Erreur

- Pour trois parutions :

$E_3^3 = 2R_1 - D_3 + D_0 = 500 - 200 = 300 \neq 125$ Erreur

$E_2^3 = 2R_1 - D_2 - D_1 = 500 - 185 - 155 = 160 \neq 90$ Erreur

$E_1^3 = 2R_1 - D_1 - D_2 = 500 - 155 - 185 = 160 \neq 195$ Erreur

$E_0^3 = 2R_1 - D_0 - D_3 = 500 - 200 = 300 \neq 590$ Erreur

Donc, la formule de Morgensztern est fausse, exception faite pour le cas E_1^2. La formule exacte a été dérivée au début de ce chapitre dans l'expression (2). On peut aussi montrer que la somme des occasions de voir définie par la formule (75) de Morgensztern n'est pas égale à la population totale :

$$\Sigma_{K=0}^{N} E_K^N = (N+1)2R_1 - \Sigma_{K=0}^{N} D_K - \Sigma_{K=0}^{N} D_{N-K}$$

Or, d'après la formule (74), nous avons :

$$\Sigma_{K=0}^{N} D_K = (N+1)R_1 - \Sigma_{K=0}^{N} R_{K+1} + \Sigma_{K=0}^{N} R_K = (N+1)R_1 - R_{N+1}$$

$$\Sigma_{K=0}^{N} D_N = (N+1)R_1 - \Sigma_{K=0}^{N} R_{N-K+1} + \Sigma_{K=0}^{N} R_{N-K} = (N+1)R_1 - R_{N+1}$$

En conséquence, la somme des occasions de voir est égale à :

$$\Sigma_{K=0}^{N} E_K^N = (N+1)\ 2R_1 - 2(N+1)\ R_1 + 2R_{N+1} = 2R_{N+1}$$

Il est aussi évident que la formule (75) de Morgensztern est symétrique, puisqu'elle vérifie $E_N^N = E_{N-K}^N$. Cette propriété est indésirable car il n'existe aucune raison a priori permettant de croire qu'il y aura autant d'individus exposés K fois que d'individus exposés N-K fois.

En fait, toutes les distributions observées des occasions de voir sont hautement dissymétriques.

Bien que Morgensztern n'ait pas réussi à trouver la relation exacte qui lie la couverture nette cumulée à la distribution des occasions de voir, il poursuit néanmoins la construction de son modèle en admettant que la couverture nette cumulée peut être représentée par une fonction exponentielle telle que :

(76) $R_N = RMAX - ab^N$ avec $0 < b < 1$

En remplaçant, dans l'équation (75), R_N par sa valeur donnée par l'équation (76) on obtient :

$$E_K^N = D\,(b^K + b^{N-K})$$

Morgensztern introduit alors deux nouveaux coefficients a et c, qui permettront de donner une forme non symétrique à la distribution des occasions de voir.

$$E_K^N = D\,(ab^K + cb^{N-K})$$

Afin que la somme de la distribution des occasions de voir soit bien égale à l'unité, Morgensztern calcule la valeur de la constante D telle que $\Sigma_{K=1}^N E_K^N = 1$. Il convient de noter que par ce choix Morgensztern exprime la distribution des occasions de voir en pourcentage de la couverture nette. Ceci entraîne la disparition de la classe des non lecteurs (i.e : E_0^N). Nous avons utilisé jusqu'ici le symbole E_K^N pour désigner la distribution des occasions de voir. Il nous faut donc maintenant introduire un nouveau symbole, T_K^N, pour désigner le pourcentage d'individus exposés à K parutions sur N parmi l'ensemble des individus exposés à au moins une parution sur N. Compte tenu de ces deux derniers changements, la formule de Morgensztern devient :

$$(77) \quad T_K^N = \frac{b-1}{b(a+1)(b^N-1)}\,(ab^K + b^{N-K+1})$$

Bien que cette expression ait été dérivée à partir de la formule (75) qui est fausse, il faut noter que les corrections qui ont été apportées, par le coefficient de dissymétrie b et le coefficient de normalisation D, suppriment les sources d'erreurs présentes dans la formule (75).

Morgensztern propose une autre justification pour l'expression (77). Il démontre en effet que l'expression (77) découle des deux hypothèses suivantes :

"-Les lecteurs d'un nombre donné de numéros sont dans un rapport constant vis à vis de l'ensemble des lecteurs d'un nombre supérieur de numéros.

- Les lecteurs d'un support se répartissent essentiellement en deux catégories, les lecteurs fidèles et les lecteurs occasionnels." (28)

Si L_K^N représente le nombre de lecteurs fidèles exposés à K parutions sur N et si O_K^N représente le nombre de lecteurs occasionnels exposés à K parutions sur N, l'hypothèse de proportionnalité s'écrira comme suit :

$$L_K^N = p\,(L_K^N + L_{K+1}^N + \ldots + L_N^N) = V_K$$

$$O_K^N = \mu\,(O_K^N + O_{K+1}^N \ldots + O_N^N) = W_K$$

L'hypothèse de proportionnalité revient donc à admettre que les fréquences d'exposition sont géométriquement distribuées dans les deux groupes de lecteurs :

$$(78) \quad \begin{cases} L_K^N = V_1 \left(\dfrac{p}{1-p}\right)(1-p)^K & \text{avec } V_1 = \Sigma_{K=1}^N L_K^N \\ O_K^N = W_1 \left(\dfrac{\mu}{1-\mu}\right)(1-\mu)^K & \text{avec } W_1 = \Sigma_{K=1}^N O_K^N \end{cases}$$

Les symboles V_1 et W_1 représentent respectivement le nombre de lecteurs fidèles et le nombre de lecteurs occasionnels parmi les individus exposés à au moins une parution parmi N.

$$R_N = V_1 + W_1$$

Morgensztern suppose que le rapport des lecteurs fidèles aux lecteurs occasionnels est une constante qui caractérise la population cible et le support étudié. Cette constante, que nous noterons a, est indépendante du nombre de parutions considéré :

$$a = V_1 / W_1$$

L'équation (78) indique que le nombre d'individus exposés à K+1 parutions est proportionnel au nombre d'individus exposés à K parutions. Cette proportion est inférieure à un pour les lecteurs occasionnels et supérieure à un pour les lecteurs fidèles.

$$L_{K+1}^N / L_K^N = 1 - p > 1$$

$$O_{K+1}^N / O_K^N = 1 - \mu < 1$$

Afin de simplifier davantage, Morgensztern suppose que ces deux proportions sont la réciproque l'une de l'autre :

$$\frac{L^N_{K+1}}{L^N_K} = \frac{O^N_K}{O^N_{K+1}} = 1 - p = \frac{1}{1 - \mu} = b$$

Le paramètre b peut donc être défini comme le rapport de deux fréquences successives dans la distribution des lecteurs fidèles. Donc b mesure le degré de fidélité parmi les lecteurs fidèles. De plus, comme b est l'inverse du rapport de deux fréquences successives dans la distribution des lecteurs occasionnels, b mesure aussi le degré d'infidélité parmi les lecteurs occasionnels.

En résumé, le modèle de Morgensztern dépend de deux paramètres :

a = le rapport des lecteurs fidèles aux lecteurs occasionnels

b = le taux de croissance de la distribution géométrique des lecteurs fidèles et l'inverse du taux de décroissance de la distribution géométrique des lecteurs occasionnels.

Une fois donné les deux paramètres a et b, on peut calculer :

- La distribution des lecteurs fidèles :

$$(79) \qquad L^N_K = \frac{a(b-1)\, b^{K-1}}{(a+1)(b^N-1)} \qquad a \geq 0, \quad b > 1 \qquad 1 \leq K \leq N$$

- La distribution des lecteurs occasionnels :

$$(80) \qquad O^N_K = \frac{(b-1)\, b^{N-K}}{(a+1)(b^N-1)} \qquad a \geq 0, \quad b > 1 \qquad 1 \leq K \leq N$$

- La distribution des fréquences pour l'ensemble des lecteurs :

$$T^N_K = L^N_K + O^N_K = \frac{(b-1)}{b(a+1)(b^N-1)} (ab^K + b^{N-K+1}) \qquad a \geq 0, \quad b > 1 \qquad 1 \leq K \leq N$$

L'exemple numérique suivant illustre le fonctionnement du modèle de Morgensztern quand il y a deux fois plus de lecteurs occasionnels que de lecteurs fidèles (i.e : a= 1/2), et quand le nombre de lecteurs occasionnels est

diminué de moitié pour chaque nouvelle exposition tandis que le nombre de lecteurs fidèles double à chaque nouvelle exposition (i.e : b=2)

Le Modèle de Morgensztern

avec a = 1/2 ; b = 2 ; N = 5

Nombre d'Expositions	1	2	3	4	5	Total
% de lecteurs fidèles	1,08	2,15	4,30	8,60	17,20	V_1 = 33,3
% de lecteurs occasionnels	34,41	17,20	8,60	4,30	2,15	W_1 = 66,7
% total de lecteurs	35,49	19,35	12,90	12,90	19,35	100,0

Bien que Morgensztern ne le mentionne pas, son modèle peut être aisément transformé afin de produire une estimation de la distribution des occasions de voir toute entière, non lecteurs compris. Dans l'expression (77) le coefficient de normalisation D est calculé de manière à ce que $\Sigma_{K=1}^{N} T_K^N = 1$. Il suffit donc de recalculer ce coefficient, en introduisant les non lecteurs, de manière à ce que la somme de la distribution complète des occasions de voir soit égale à l'unité. C'est à dire :

$$\Sigma_{K=0}^{N} E_K^N = 1 \qquad \text{avec} \qquad E_K^N = D(ab^K + b^{N-K+1})$$

Il vient : $\Sigma_{K=0}^{N}(ab^K + b^{N-K+1}) = a\Sigma_{K=0}^{N} b^K + \Sigma_{K=0}^{N} b^{N-K+1} = \dfrac{(a+b)(b^{N+1} - 1)}{b - 1}$

En conséquence, nous avons : $D = \dfrac{b - 1}{(a+b)(b^{N+1} - 1)}$

Le modèle de Morgensztern, étendu à la totalité de la distribution des occasions de voir est donc donné par la formule suivante :

$$(81) \qquad E_K^N = \frac{b - 1}{(a+b)\ (b^{N+1}-1)} \quad (ab^K + b^{N-K+1}) \qquad a \geq 0\ ,\ b > 1 \quad 0 \leq K \leq N$$

Toutefois, la signification du paramètre "a" a maintenant changé. Pour le montrer, nous allons établir les nouvelles formules relatives aux lecteurs fidèles et aux lecteurs occasionnels, en séparant en deux parties la formule (81)

- Distribution des occasions de voir parmi les lecteurs fidèles :

$$(82) \qquad L_K^N = \frac{a(b-1)b^K}{(a+b)(b^{N+1}-1)} \qquad \begin{array}{l} a \geq 0 \ , \ b \geq 1 \\ 0 \leq K \leq N \end{array}$$

- Distribution des occasions de voir parmi les lecteurs occasionnels :

$$(83) \qquad O_K^N = \frac{(b-1)b^{N-K+1}}{(a+b)(b^{N+1}-1)} \qquad \begin{array}{l} a \geq 0 \ , \ b \geq 1 \\ 0 \leq K \leq N \end{array}$$

Ces deux formules montrent que b est toujours égal au rapport de deux fréquences successives parmi les lecteurs fidèles, et que 1/b reste égal au rapport de deux fréquences successives parmi les lecteurs occasionnels. Cependant, le rapport du nombre de lecteurs fidèles au nombre de lecteurs occasionnels est maintenant donné par :

$$\frac{V1}{W_1} = \frac{a/(a+b)}{b/(a+b)} = \frac{a}{b}$$

L'exemple numérique suivant illustre le fonctionnement du modèle de Morgensztern, étendu à la totalité de la distribution des occasions de voir, dans le cas des cinq parutions.

Modèle de Morgensztern étendu

a = 1/2 ; b = 2 ; N = 5

Nombre d'Expositions	0	1	2	3	4	5
% de Lecteurs Fidèles	0,32	0,63	1,27	2,54	5,08	10,16
% de Lecteurs Occasionnels	40,63	20,32	10,16	5,08	2,54	1,27
Population Totale	40,95	20,95	11,43	7,62	7,62	11,43

Maintenant que nous avons étendu le modèle de Morgensztern à la distribution complète des occasions de voir, nous pouvons établir la formule de la couverture nette cumulée :

$$R_N = E_K^N / T_K^N$$

(84)
$$R_N = \frac{b(a+1)(b^N-1)}{(a+b)(b^{N+1}-1)}$$

Cette formule qui n'a pas été établie par Morgensztern, nous permet d'estimer directement le paramètre b à partir de l'audience cumulée des deux premières parutions :

$$R_1 = \frac{b(a+1)(b-1)}{(a+b)(b^2-1)}$$

$$R_2 = \frac{b(a+1)(b^2-1)}{(a+b)(b^3-1)}$$

En divisant R_2 par R_1 et en éliminat le paramètre a, nous obtenons :

$$\frac{R_2}{R_1} = \frac{(b^2-1)^2}{(b^3-1)(b-1)} = \frac{b^3+b^2-b-1}{b^3-1}$$

En réécrivant cette expression, il vient :

(85) $$(R_1-R_2)\, b^3 + R_1\, b^2 - R_1\, b + R_2 - R_1 = 0$$

L'équation (85) est une cubique qui, selon les valeurs prises par R_1 et R_2, aura soit une racine réelle et deux racines imaginaires conjuguées, soit trois racines réelles dont deux au moins seront égales, soit enfin trois racines réelles quelconques. Cependant, le modèle de Morgenztern ne s'applique que si b est supérieur ou égal à l'unité. Or, rien ne prouve qu'au moins une des racines de l'équation (85) sera supérieure à l'unité. Ce qui signifie que le modèle de Morgensztern ne peut pas être ajusté pour des valeurs quelconques de R_1 et R_2. Ceci, bien entendu, est une grave faiblesse du modèle. En fait, parmi nos dix magazines, un seul remplit les conditions d'applicabilité du

modèle de Morgensztern.

Afin de préciser davantage les valeurs admissibles de R_1 et R_2, notons que l'équation (85) doit changer de signe dans l'interval $]1, \infty[$. En calculant la dérivée première de la cubique, cette condition peut s'écrire comme suit :

$$3(R_1-R_2)\ b^2 + 2R_1\ b-R_1 = 0 \qquad \text{avec} \qquad 1 < b < \infty$$

Cette équation du second degré aura des racines réelles si :

$$4\ R_1^2 + 12\ R_1\ (R_1 - R_2) \geq 0$$

Ce qui équivaut à la condition suivante :

$$(86) \qquad R_2 \leq \frac{4}{3}\ R_1$$

Nous avons donc découvert que le modèle de Morgensztern n'est applicable que si et seulement si R_2 est inférieur ou égal aux quatre tiers de R_1. Quand cette condition est remplie, le second paramètre du modèle peut être estimé comme suit :

$$(87) \quad \hat{a} = \hat{b}\ \{1-R_1(\hat{b}+1)\}\ /\ \{R_1(\hat{b}+1)-\hat{b}\} = \hat{b}\ \{\hat{b}^2-1-R_2(\hat{b}^3-1)\}\ /\ \{R_2(\hat{b}^3-1)-\hat{b}^3+\hat{b}\}$$

La table suivante décrit les résultats du test d'applicabilité du modèle de Morgensztern pour nos dix magazines.

Test d'Applicabilité du Modèle de Morgensztern

Magazines	R_1	R_2	R_2 maximum pour applicabilité	Conclusion
Life 1953	0,221	0,324	> 0,295	Le modèle ne s'applique pas
Post 1953	0,118	0,178	> 0,157	Le modèle ne s'applique pas
Look 1953	0,151	0,239	> 0,201	Le modèle ne s'applique pas
L.H.J. 1953	0,096	0,144	> 0,128	Le modèle ne s'applique pas
Week 1953	0,192	0,259	> 0,256	Le modèle ne s'applique pas
Life 1958	0,312	0,426	> 0,416	Le modèle ne s'applique pas
Post 1958	0,194	0,274	> 0,259	Le modèle ne s'applique pas
Look 1958	0,238	0,344	> 0,317	Le modèle ne s'applique pas
L.H.J. 1958	0,194	0,273	> 0,259	Le modèle ne s'applique pas
Digest 1958	0,360	0,465	< 0,480	Le modèle s'applique

Il apparaît donc que seul le magazine Reader's Digest remplit la condition d'application du modèle de Morgensztern. L'équation (85) est égale à :

$$- 0{,}105\ b^3 + 0{,}36\ b^2 - 0{,}36\ b + 0{,}105 = 0$$

Cette équation admet une racine réelle supérieure à l'unité $b = 1{,}9024$. La paramètre a, estimé par l'équation (87) est égal à : $\hat{a} = 0{,}0994$. La qualité de l'ajustement est peu satisfaisante puisque l'erreur type de prédiction est de 0,053. La table suivante compare la couverture nette et la distribution des occasions de voir estimées à la couverture nette et à la distribution des occasions de voir observées.

Le Modèle d'Accumulation de Morgensztern

Reader's Digest - 1958

	Couverture Nette Cumulée		Distribution des Occasions de Voir		
Parutions	Observées	Estimées	Expositions	Observées	Estimées
1	36,0	36,0	0	37,8	45,6
2	46,5	46,5	1	15,0	24,1
3	52,5	50,8	2	9,5	12,8
4	56,6	52,9	3	6,9	7,0
5	59,7	53,9	4	8,5	4,1
6	62,2	54,4	5	6,8	3,1

Erreur type = 0,053 ; $\hat{a}$ = 0,099 ; $\hat{b}$ = 1,902 ; Chi Deux = 22,2

Compte tenu des divers types d'erreurs affectant la mesure de R_1 et de R_2, on pourrait être tenté d'imputer à ces erreurs de mesure le fait que la condition d'application du modèle n'a pas été vérifiée dans 9 cas sur 10. En suivant cette ligne de raisonnement, le modèle ne serait pas en cause, seules les données seraient fautives. Toutefois, nous pouvons aussi estimer les paramètres a et b, à partir de la couverture nette cumulée observée pour six parutions, par la méthode des moindres carrés. Il suffit de déterminer les valeurs de a et b qui minimisent la somme des carrés des écarts entre la couverture nette observée et la couverture nette estimée pour les six parutions. C'est ce que nous avons fait et les deux tables suivantes permettent de juger de la validité de l'argument tendant à imputer l'échec du modèle aux erreurs de mesure de R_1 et R_2.

La première table décrit l'ajustement, par la méthode des moindres carrés, du modèle de Morgensztern à la couverture nette cumulée. Pour chaque magazine la qualité de l'ajustement est peu satisfaisante et les erreurs sont systématiques. Le modèle ne reproduit pas correctement le taux d'accumulation observé. Il surestime la couverture nette des deux premières parutions et sous-estime celle des deux dernières parutions. Il convient de remarquer que l'erreur type moyenne est de 0,0463 pour les neuf premiers magazines, alors qu'elle n'est que de 0,0202 pour Reader's Digest, le seul magazine qui remplisse la condition d'application du modèle.

Le Modèle d'Accumulation de Morgensztern
estimé par la méthode des moindres carrés
Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
1	22,1	30,6	11,8	19,8	15,1	25,3	9,6	16,0	19,2	24,1
2	32,4	38,9	17,8	22,7	23,9	31,1	14,4	18,5	25,9	29,4
3	39,1	41,9	21,8	23,5	30,0	32,9	17,7	18,9	29,6	30,9
4	44,0	43,2	25,0	23,7	34,8	33,5	20,2	19,0	32,0	31,5
5	47,7	43,7	27,5	23,8	38,6	33,7	22,2	19,1	33,8	31,6
6	50,6	44,0	29,6	23,9	41,8	33,8	23,9	19,1	35,3	31,7
$\hat{a}$	0,0004		0,005		0,0001		0,0006		0,0024	
$\hat{b}$	2,266		4,209		2,959		5,248		3,162	
Erreur Type	0,0553		0,0469		0,0651		0,0396		0,0307	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
1	31,2	36,2	19,4	26,5	23,8	31,8	19,4	26,0	36,0	37,7
2	42,6	47,1	27,4	32,9	34,4	40,6	27,3	32,2	46,5	49,2
3	49,4	51,7	32,6	34,9	41,2	44,0	32,2	34,1	52,5	54,3
4	54,1	54,0	36,5	35,6	46,0	45,4	35,6	34,8	56,6	56,9
5	57,7	55,2	39,6	35,9	49,7	46,0	38,3	35,0	59,7	58,4
6	60,5	55,9	42,1	36,0	52,7	46,3	40,4	35,1	62,2	59,2
$\hat{a}$	0,0006		0,0013		0,0053		0,0110		0,0031	
$\hat{b}$	1.763		2.781		2.159		2.867		1.658	
Erreur Type	0,0360		0,0480		0,0529		0,0432		0,0202	

Le Modèle d'Accumulation de Morgensztern
estimé par la méthode des moindres carrés
Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	56,0	70,4	76,2	58,2	66,2	76,1	80,9	64,7	68,3
1	17,4	24,7	12,6	18,1	19,2	22,4	10,2	15,4	9,0	21,6
2	12,0	10,9	6,8	4,3	9,8	7,6	4,5	2,9	4,5	6,8
3	8,0	4,8	3,0	1,0	5,0	2,6	4,0	0,6	6,0	2,2
4	3,0	2,1	1,5	0,2	3,0	0,9	1,5	0,1	6,0	0,7
5	4,2	1,0	1,5	0,1	2,2	0,3	1,8	0,1	3,6	0,2
6	6,0	0,5	4,2	0,1	2,6	0,0	1,9	0,0	6,2	0,2
Chi Deux	13,3		11,7		8,9		11,4		35,4	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	44,1	57,9	63,9	47,3	53,7	59,6	64,9	37,8	40,8
1	16,7	25,0	15,1	23,1	17,8	24,9	12,4	22,6	15,0	24,6
2	11,0	14,2	8,2	8,3	11,3	11,6	8,6	7,9	9,5	14,9
3	9,4	8,0	4,9	3,0	7,3	5,4	4,9	2,8	6,9	9,0
4	6,7	4,6	5,2	1,1	6,4	2,5	5,2	1,0	8,5	5,4
5	6,6	2,6	2,5	0,4	4,4	1,2	4,0	0,4	6,8	3,3
6	10,1	1,5	6,2	0,2	5,5	0,7	5,3	0,4	15,5	2,0
Chi Deux	16,2		16,5		13,1		21,1		24,6	

A l'examen de cette première table, il apparait donc que l'argument tendant à imputer l'échec du modèle aux erreurs de R_1 et R_2 n'est pas validé. En effet, quand on utilise la couverture nette des six parutions plutôt que des deux premières, l'ajustement du modèle est mauvais et les erreurs sont systématiques pour les magazines qui ne remplissent pas la condition d'application.

De plus, l'examen de la distribution estimée des occasions de voir montre que les mauvais résultats du modèle sont dus à son incapacité à reproduire correctement la forme en U des distributions observées. Dans le cas du magazine Reader's Digest, le Chi Deux obtenu sur la base des six parutions est supérieur au Chi Deux obtenu sur la base de six parutions est supérieur au Chi Deux obtenu sur la base des deux premières parutions. La conclusion qui s'impose est donc que l'argument des erreurs de mesure dans R_1 et R_2 n'est pas valide. Quand la condition d'application n'est pas remplie, il est inutile d'essayer de passer outre.

Le dernier modèle d'accumulation à double hétérogénéité que nous allons examiner représente une extension logique du modèle Bêta Binomial. La densité de probabilité Bêta étant continue, elle donne un poids nul aux probabilités d'expositions $p = 0$ et $p = 1$. Cela signifie qu'aucun individu de la population cible n'est considéré comme étant non lecteur absolu ou un lecteur parfaitement fidèle. Il est apparu à l'auteur qu'un meilleur modèle pouvait être construit en ajoutant au segment des lecteurs probables un segment de non lecteurs absolus et un segment de lecteurs parfaitement fidèles.

Le Modèle Bêta Binomial Segmenté

Nous étudierons d'abord une version à deux segments, puis une version à trois segments. Nous admettrons donc, pour commencer, que la population cible se compose d'un segment homogène de lecteurs absolument fidèles dont la probabilité d'exposition est égale à un et qui représente W pour cent de la population, et d'un segment hétérogène de lecteurs probables qui représente 1 - W pour cent de la population. Les probabilités d'expositions des lecteurs probables suivent une distribution Bêta de paramètres A et B.

Segment	Probabilité	Pourcentage de la population
- Lecteurs absolument fidèles	$p = 1$	$f(1) = W$
- Lecteurs probables	$0 < p < 1$	$f(p) = (1-W)\frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} p^{A-1}(1-p)^{B-1}$

La distribution des occasions de voir est similaire à celle du modèle Bêta Binomial, excepté qu'elle ne s'applique qu'à 1-W pour cent de la population cible.

$$(88) \quad E_K^N = \begin{cases} W + (1-W)\dfrac{(A+K-1)\ \ldots.(A)}{(A+B+N-1\ \ldots.(A+B)} & \text{si } K = N \\[2ex] (1-W)\binom{N}{K}\dfrac{(A+K-1)\ \ldots.(A)(B+N-K-1)\ \ldots.\ (B)}{(A+B+N-1)\ \ldots\ldots\ldots\ldots\ (A+B)} & \text{si } K < N \end{cases}$$

$$0 \leq K \leq N \quad , \quad \text{entiers} \quad ; A, B > 0 \ , \text{réels}$$

A partir de l'expression (88), la formule de la couverture nette est aisément dérivée :

$$(89) \quad R_N = 1 + (W-1)\frac{(B+N-1)\ \ldots.\ (B)}{(A+B+N-1)\ \ldots.\ (A+B)}$$

Cette version à deux segments ne dépend que de trois paramètres, A, B, W. Il suffira donc de connaître la couverture nette cumulée des trois premières parutions, pour estimer ces paramètres.

A partir de l'expression (88), nous pouvons exprimer A et B en fonction de E_0^1 , E_0^2 , E_0^3 , et ensuite éliminer le paramètre W.

$$\begin{cases} E_0^2 \ / \ E_0^1 = (B+1) \ / \ (A+B+1) \\[2ex] E_0^3 \ / \ E_0^2 = (B+2) \ / \ (A+B+2) \end{cases}$$

En résolvant des deux équations pour les paramètres A et B, nous obtenons :

$$(90) \quad \begin{cases} \hat{B} = \dfrac{2(E_0^2)^2 - E_0^3(E_0^1 + E_0^2)}{E_0^1\ E_0^3 - (E_0^2)^2} \\[4ex] \hat{A} = \dfrac{(\hat{B}+1)\ (E_0^1 - E_0^2)}{E_0^2} \end{cases}$$

A partir des valeurs estimées de A et B et de l'audience moyenne d'une parution, E_1^1, nous obtenons l'estimation de W :

$$(91) \qquad \hat{W} = \frac{E_1^1\,(\hat{A}+\hat{B}) - \hat{A}}{\hat{B}}$$

Puisque les trois paramètres doivent être positifs, les conditions d'application du modèle sont les suivantes :

$$(92) \qquad \begin{cases} E_0^3 < E_0^2 < E_0^1 \\ \hat{A} < E_1^1\,(\hat{A}+\hat{B}) \end{cases}$$

La première condition est toujours vérifiée puisqu'elle implique seulement que la couverture nette s'accroisse d'une parution à une autre. La seconde condition, qui assure la positivité du paramètre W, est vérifiée par 9 magazines sur 10. Elle n'est pas vérifiée pour le magazine This Week. Quand le cas se produit, il suffit de poser W = 0 pour retrouver le modèle Bêta Binomial simple.

A l'aide des expressions (90) et (91) nous avons ajusté le modèle Bêta Binomial à deux segments pour nos dix magazines. Puis, utilisant les expressions (89) et (88) nous avons calculé la couverture nette estimée et la distribution des occasions de voir estimée. Les deux tables suivantes comparent les estimations du modèle aux données observées.

Nous observons que le modèle Bêta Binomial à deux segments reproduit extrêmement bien la couverture nette cumulée. L'erreur type moyenne de prédiction est seulement de 0,002. C'est de loin le meilleur résultat obtenu jusqu'ici. L'amélioration, par rapport au modèle d'Agostini et au modèle Bêta Binomial, est très importante. De plus, nous n'observons aucun biais systématique puisque le modèle surestime la couverture nette dans 6 cas et la sousestime dans 4 cas. Finalement, nous remarquons que les estimations des paramètres A et B n'ont pas changé beaucoup par rapport à celles du modèle Bêta Binomial. Dans tous les cas les deux paramètres ont des valeurs légèrement plus élevées, ce qui entraîne un taux d'accumulation supérieur qui corrige la sousestimation observée pour le modèle Bêta Binomial simple. Le pourcentage de lecteurs absolument fidèles reste faible pour tous les magazines. Le pourcentage le plus élevé, 7,9 %; apparaît pour le magazine Reader's Digest. Ce résultat est logique compte renu du fait que ce magazine est vendu principalement par souscription.

Toutefois, les améliorations apportées par le modèle ont un prix, la connaissance de la couverture nette cumulée de la troisième parution qui n'était pas nécessaire pour le modèle Bêta Binomial simple.

Le Modèle Bêta Binomial à Deux Segments

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
4	44,0	43,9	25,0	24,8	34,8	34,6	20,2	20,2	32,0	32,5
5	47,7	47,6	27,5	27,1	38,6	38,2	22,2	22,2	33,8	34,6
6	50,6	50,6	29,6	29,0	41,8	41,1	23,9	23,8	35,3	36,3
$\hat{a}$	0,396		0,171		0,354		0,141		0,146	
$\hat{b}$	1,598		1,342		2,058		1,510		0,614	
% de lecteurs fidèles	2,8		0,6		0,5		1,2		0,0	
Erreur Type	0,0006		0,0044		0,0047		0,0005		0,0070	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
4	54,1	54,1	36,5	36,4	46,0	46,1	35,6	35,7	56,6	56,5
5	57,7	57,5	39,6	39,3	49,7	49,8	38,3	38,4	59,7	59,5
6	60,5	60,2	42,1	41,7	52,7	52,7	40,4	40,5	62,2	61,9
$\hat{a}$	0,416		0,257		0,407		0,216		0,354	
$\hat{b}$	1,093		1,335		1,517		0,986		0,806	
% de lecteurs fidèles	5,0		3,9		3,4		1,8		7,9	
Erreur Type	0,0018		0,0031		0,0005		0,0008		0,0022	

Le Modèle Bêta Binomial à Deux Segments

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	49,4	70,4	71,0	58,2	58,9	76,1	76,2	64,7	63,7
1	17,4	17,8	12,6	11,5	19,2	17,7	10,2	9,9	9,0	10,0
2	12,0	11,1	6,8	6,3	9,8	9,9	4,5	5,1	4,5	6,2
3	8,0	7,8	3,0	4,2	5,0	6,1	4,0	3,2	6,0	4,9
4	3,0	5,5	1,5	3,0	3,0	3,8	1,5	2,2	6,0	4,4
5	4,2	3,8	1,5	2,1	2,2	2,2	1,8	1,5	3,6	4,5
6	6,0	2,7	4,2	1,9	2,6	1,4	1,9	1,9	6,2	6,3
Chi Deux	4,7		4,3		1,7		1,2		1,6	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	39,8	57,9	58,3	47,3	47,3	59,6	59,5	37,8	38,1
1	16,7	16,3	15,1	14,3	17,8	17,7	12,4	12,9	15,0	14,0
2	11,0	11,0	8,2	8,4	11,3	11,3	8,6	7,9	9,5	9,9
3	9,4	8,9	4,9	5,9	7,3	8,1	4,9	5,8	6,9	8,2
4	6,7	6,7	5,2	4,3	6,4	5,9	5,2	4,7	8,5	7,4
5	6,6	6,3	2,5	3,1	4,4	4,1	4,0	4,0	6,8	7,1
6	10,1	9,9	6,2	5,7	5,5	5,6	5,3	5,2	15,5	15,3
Chi Deux	1,3		1,4		0,8		0,6		0,5	

La distribution des occasions de voir estimée par le modèle Bêta Binomial à deux segments est aussi extrêmement satisfaisante. Le Chi Deux moyen est seulement de 1,8. Ce résultat est meilleur que celui obtenu par le modèle Bêta Binomial simple. Pour 8 magazines sur 10 le test du Chi Deux indique que les fréquences estimées sont identiques aux fréquences observées, au seuil de confiance de 95%.

Avant de quitter le domaine des modèles d'audience stationnaires, nous allons étudier un modèle à double hétérogénéité qui introduit des non lecteurs et des lecteurs absolus aussi bien que des lecteurs probables. Nous admettrons donc que la population cible se compose des trois segments suivants :

Segments	Probabilité	Pourcentage de la Population
- Lecteurs absolus	1	$f(1) = W_1$
- Lecteurs probables	$0 < p < 1$	$f(p) = (1-W_1-W_0)\,\dfrac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\,p^{A-1}(1-p)^{B-1}$
- Non lecteurs absolus	0	$f(0) = W_0$

Dans ces conditions, la distribution des occasions de voir est donnée par :

$$(93)\quad E_K^N = \begin{cases} W_1 + (1-W_1-W_0)\dfrac{(A+K-1)\ \ldots\ (A)}{(A+B+N-1)\ \ldots\ (A+B)} & \text{si } K = N \\[2ex] (1-W_1-W_0)\dbinom{N}{K}\dfrac{(A+K-1)\ \ldots\ (A)(B+N-K-1)\ \ldots\ (B)}{(A+B+N-1)\ \ldots\ldots\ldots\ldots\ (A+B)} & \text{si } 0 < K < N \\[2ex] W_0+(1-W_1-W_0)\dfrac{(B+N-1)\ \ldots\ (B)}{(A+B+N-1)\ \ldots\ (A+B)} & \text{si } K = 0 \end{cases}$$

Si nous appelons W_p la proportion de lecteurs probables, $W_p = 1 - W_1 - W_0$

La couverture nette cumulée s'exprime comme suit :

$$(94)\quad R_N = 1 - W_0 - W_p * \frac{(B+N-1)\ \ldots\ (B)}{(A+B+N-1)\ \ldots\ (A+B)}$$

Le modèle Bêta Binomial à 3 segments dépend de 4 paramètres, nous devons donc connaître la couverture nette cumulée de 4 parutions successives pour estimer les valeurs de A, B, W_0, W_1. A partir de l'expression (93), nous pouvons écrire les rapports successifs entre les individus non exposés dans le segment des lecteurs probables :

$$\begin{cases} (E_0^2 - W_0) \,/\, (E_0^1 - W_0) = (B+1) \,/\, (A+B+1) \\ (E_0^3 - W_0) \,/\, (E_0^2 - W_0) = (B+2) \,/\, (A+B+2) \\ (E_0^4 - W_0) \,/\, (E_0^3 - W_0) = (B+3) \,/\, (A+B+3) \end{cases}$$

Ce système peut être résolu pour les valeurs des trois paramètres inconnus.

$$(95) \quad \begin{cases} \hat{B} = \dfrac{E_0^1(E_0^4-E_0^3) + E_0^2(2E_0^2-4E_0^3) + E_0^3(3E_0^3-E_0^4)}{E_0^1(E_0^3-E_0^4) + E_0^2(E_0^3+E_0^4-E_0^2) - (E_0^3)^2} \\[2ex] \hat{A} = \dfrac{\hat{B}(E_0^3+E_0^1-2E_0^2) + 2E_0^3 + E_0^1 - 3E_0^2}{E_0^2 - E_0^3} \\[2ex] \hat{W}_0 = \dfrac{E_0^2(\hat{A}+\hat{B}+1) - E_0^1(\hat{B}+1)}{\hat{A}} \end{cases}$$

Le paramètre restant, W_1, est obtenu en introduisant l'audience moyenne d'une parution, E_1^1 :

$$(96) \qquad \hat{W}_1 = (E_1^1(\hat{A}+\hat{B}) - (1-W_0)\,\hat{A}) \,/\, \hat{B}$$

Puisque les quatre paramètres doivent être positifs, les conditions d'application du modèle sont les suivantes :

1- Le paramètre B sera positif lorsque le numérateur et le dénominateur qui le compose seront simultanément positifs ou négatifs :

Le numérateur de B est positif si : $3(E_0^3)^2+2(E_0^2)^2+E_0^1E_0^4 > E_0^3(E_0^1+4E_0^2+E_0^4)$

Le dénominateur de B est positif si : $E_0^3E_0^1+E_0^3E_0^2+E_0^4E_0^2 > (E_0^3)^2 + E_0^4E_0^1$

2- Si B est positif, A sera positif quand :

$$E_0^2(2B+3) < E_0^3(B+2) + E_0^1(B+1)$$

3- Si A et B sont positifs, W_0 sera positif et inférieur à l'unité si :

$$E_0^1(B+1) < E_0^2(A+B+1) < A+E_0^1(B+1)$$

4- Enfin, si A et B sont positifs et si W_0 est compris entre zéro et un, alors W_1 sera positif et inférieur à $1-W_0$ quand :

$$(1-W_0)\ A < E_1^1(A+B) < (1-W_0)(A+B)$$

La condition d'application numéro un est facilement satisfaite par les données. Pae contre, les trois autres conditions sont plus sévères. La table suivante montre qu'un seul magazine parmi les 10 remplit simultanément les quatre conditions d'application du modèle Bêta Binomial à 3 segments.

Test d'Applicabilité du Modèle Bêta Binomial à 3 Segments

Magazines	$B > 0$	$A > 0$	$0 < W_0 < 1$	$0 < W_1 < 1-W_0$	Conclusion : Modèle applicable
Life 1953	OUI	OUI	NON	-	2 segments Bêta Binomial
Post 1953	OUI	OUI	NON	-	2 segments Bêta Binomial
Look 1953	OUI	NON	-	-	2 segments Bêta Binomial
LHJ 1953	OUI	OUI	NON	-	2 segments Bêta Binomial
Week 1953	OUI	OUI	OUI	OUI	3 segments Bêta Binomial
Life 1958	OUI	OUI	NON	-	2 segments Bêta Binomial
Post 1958	OUI	OUI	NON	-	2 segments Bêta Binomial
Look 1958	OUI	OUI	NON	-	2 segments Bêta Binomial
LHJ 1958	OUI	OUI	NON	-	2 segments Bêta Binomial
Digest 1958	OUI	OUI	NON	-	2 segments Bêta Binomial

This week est le seul magazine, parmi ceux étudiés, à posséder un segment important de non lecteurs absolus. L'ajustement du modèle à 3 segments est particulièrement bon. L'erreur type de prédiction est seulement de 0,0023. La distribution des occasions de voir est extrêmement bien reproduite avec une valeur du Chi Deux de 1,4. La table suivante compare les estimations du modèle Bêta Binomial à 3 segments avec les données observées.

Le Modèle Bêta Binomial à 3 segments

Magazine This Week - 1953

Couverture Nette			Distribution des Occasions de Voir		
Parutions	Observées	Estimées	Expositions	Observées	Estimées
5	33,8	33,7	0	64,7	65,0
6	35,3	35,0	1	9,0	7,7
			2	4,5	6,3
$\hat{A}$	0,632		3	6,0	5,5
$\hat{B}$	1,013		4	6,0	5,0
$\hat{W}_0$	52,8%		5	3,6	4,6
$\hat{W}_1$	1,7%		6	6,2	5,9
Erreur Type	0,0023		Chi Deux	1,4	

En résumé, le modèle Bêta Binomial segmenté obtient des résultats supérieurs à ceux de tous les autres modèles d'accumulation. Toutefois, il requiert la connaissance de la couverture nette de 3 ou 4 parutions, selon la version utilisée. De plus, les conditions d'application du modèle sont plus sévères que celles requises par les modèles à hétérogénéité individuelle.

Nous avons terminé notre étude des modèles d'ordre zéro. Nous allons maintenant considérer un modèle qui supprime l'hypothèse de stationnarité des probabilités d'exposition.

325. Modèle d'Accumulation d'ordre un

Le modèle Binomial Markovien

Ce modèle abandonne l'hypothèse de stationnarité des probabilités d'exposition d'une parution à une autre. Nous faisons désormais l'hypothèse que la probabilité d'exposition à une parution dépend de l'exposition ou de la non exposition à la parution précédente. Cependant, afin de ne pas tomber dans une complexité insurmontable, nous adopterons l'hypothèse d'homogénéité de la population et nous conserverons l'hypothèse de la constance de l'audience d'une parution à une autre. Nous pouvons donc résumer les hypothèses du modèle comme suit :

a- La population cible est homogène. Tous les individus exposés à la dernière parution possèdent la même probabilité d'exposition à la parution suivante. Tous les individus non exposés à la dernière parution possèdent la même probabilité d'exposition à la parution suivante. Mais, bien entendu, ces deux probabilités ne sont pas égales.

b- La probabilité d'exposition dépend de l'exposition à la parution précédente. Elle est plus élevée pour les individus qui ont été exposés et plus faible pour les individus qui n'ont pas été exposés.

c- L'audience d'une parution reste constante d'une parution à une autre.

Avant d'aller plus loin, il convient de noter que ce modèle, comme son nom le suggère, n'est rien d'autre qu'une extension pure et simple du modèle binomial au cas de non stationnarité des probabilités d'exposition. De plus, la dépendance par rapport aux expositions précédentes ne s'étend qu'à la dernière parution. On ne prend pas en considération l'histoire complète de toutes les expositions précédentes.

Nous utiliserons les notations suivantes :

P(1/1 = d = Probabilité d'exposition à la prochaine parution des individus exposés à la parution précédente.

P(0/1) = Probabilité de non exposition à la prochaine parution des individus exposés à la parution précédente.

P(1/0) = Probabilité d'exposition à la prochaine parution des individus non exposés à la parution précédente

P(0/0) = Probabilité de non exposition à la prochaine parution des individus non exposés à la parution précédente.

p = Probabilité inconditionnelle d'exposition à une parution quelconque.

q = Probabilité inconditionnelle de non exposition à une parution quelconque.

Par définition, nous avons les trois égalités suivantes :

$$(97) \quad \begin{cases} p + q = 1 \\ P(1/1) + P(0/1) = 1 \\ P(1/0) + P(0/0) = 1 \end{cases}$$

Puisque nous désirons que p reste constant d'une parution à une autre, nous devons avoir :

$$(98) \quad \begin{cases} p*P(1/1) + q*P(1/0) = p \\ p*P(0/1) + q*P(0/0) = q \end{cases}$$

Ainsi, le modèle est entièrement spécifié lorsque d et p sont connus. Il est facile d'estimer d et p à partir de l'audience cumulée des deux premières parutions :

$$(99) \quad \hat{p} = \text{Audience d'une parution / Population} = R_1 \quad (\text{en } \%)$$

La probabilité conditionnelle d'exposition en cas d'exposition à la parution précédente est égale au rapport de l'audience dupliquée de deux parutions successives à l'audience moyenne d'une parution :

$$(100) \quad \hat{d} = \text{Audience dupliquée / Audience d'une parution} = (2R_1 - R_2)/R_1$$

Les deux paramètres d et p étant estimés, nous pouvons spécifier la matrice de transition en utilisant les équations (97) et (98).

Matrice de Transition

		Prochaine Parution	
		Exposition	Non Exposition
Dernière Parution	Exposition	$P(1/1) = d$	$P(0/1) = 1-d$
	Non Exposition	$P(1/0) = (1-d)p/q$	$P(0/0) = (1-2p + dp) / q$

Il faut noter que, lorsque d = p, c'est à dire lorsque la proportion d'individus exposés est la même parmi les individus précédemment exposés que parmi l'ensemble des individus, le modèle se réduit au modèle binomial simple. Quand d est supérieur à p, c'est à dire lorsque le fait d'avoir été exposé rend plus probable une nouvelle exposition, l'audience s'accumule plus lentement qu'elle ne s'accumulerait au pur hasard, ce qui correspond à toutes les observations empiriques. Puisque d doit être compris entre zéro et un, la condition d'application du modèle est la suivante :

$$(101) \qquad R_1 < R_2 < 2R_1$$

Les propriétés statistiques du modèle Binomial Markovien ont été étudiées par J. Klotz (49) et seront analysés en détail dans le chapitre 7 de la thèse. Pour qu'un individu soit exposé à N parutions parmi N, il faut qu'il soit exposé à la première parution, ce qui se produit avec la probabilité p, et qu'il reste exposé à chacune des N-1 parutions suivantes, ce qui se produit avec la probabilité $(d)^{N-1}$.

On obtient donc :

$$P(K=N) = p(d)^{N-1} = R_1\left[\frac{2R_1 - R_2}{R_1}\right]^{N-1}$$

Puisque la population est homogène, la distribution des lecteurs fidèles sera donnée par :

$$(102) \qquad E_N^N = POP * R_1\left[\frac{2R_1 - R_2}{R_1}\right]^{N-1}$$

Pour qu'un individu reste non exposé après N parutions, il faut qu'il ne soit pas exposé à la première parution, ce qui se produit avec la probabilité q, et qu'il reste non exposé aux N-1 parutions suivantes, ce qui se produit avec la probabilité $(P(0/0))^{N-1}$. On obtient donc :

$$P(K=0) = q\left[\frac{1-2p+dp}{q}\right]^{N-1} = (1-R_1)\left[\frac{1-R_2}{1-R_1}\right]^{N-1}$$

Donc, l'équation de la couverture nette est donnée par :

$$(103) \qquad R_N = POP * \left[1-(1-R_1)\left[\frac{1-R_2}{1-R_1}\right]^{N-1}\right]$$

Nous n'avons pas pu dériver une équation directe pour la distribution des occasions de voir. Cependant, nous pouvons la calculer de façon indirecte soit à partir des expressions (102) ou (103), soit en recourant à la formule de conversion présentée au début de ce chapitre. Les formules de récurrence suivantes s'avèrent les plus commodes pour les calculs numériques.

Soit P_N (... 1/1) la probabilité qu'un individu soit exposé à la N ième parution quand il a été exposé à la N-1 ième parution. En vertu des probabilités de la matrice de transition, nous obtenons les équations de récurrence suivantes :

$$\begin{cases} P_N\ (\ \ldots\ 1/1) = P_{N-1}\ (\ldots\ 1) * (2R_1-R_2)/R_1 \\ P_N\ (\ \ldots\ 1/0) = P_{N-1}\ (\ldots\ 1) * (R_2-R_1)/R_1 \\ P_N\ (\ \ldots\ 0/1) = P_{N-1}\ (\ldots\ 0) * (R_2-R_1)/(1-R_1) \\ P_N\ (\ \ldots\ 0/0\ = P_{N-1}\ (\ldots\ 0) * (1-R_2)/(1-R_1) \end{cases}$$

Comme nous l'avons fait pour le modèle Bêta Binomial, nous pouvons calculer, à partir de la matrice de transition, le coefficient de corrélation entre les audiences de deux parutions successives. Nous obtenons :

$$(104) \qquad \emptyset = \frac{d-p}{q} = \frac{2R_1 - R_2 - (R_1)^2}{R_1\ (1-R_1)}$$

Nous observons que, lorsque d = p c'est à dire quand le modèle Binomial Markovien se réduit au modèle Binomial, la corrélation entre les audiences de deux parutions devient nulle. Quand d tend vers l'unité, c'est à dire quand tout individu exposé à une parution le sera nécessairement à la prochaine, la corrélation tend aussi vers l'unité et le taux d'accumulation de l'audience devient nul.

Les deux tables suivantes décrivent les performances du modèle Binomial Markovien en matière de prédiction de la couverture nette et de la distribution des occasions de voir. Les estimations sont basées sur l'audience cumulée des deux premières parutions de nos dix magazines habituels.

Le Modèle d'Accumulation Binomial Markovien

Estimation de la Couverture Nette

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	39,1	41,3	21,8	23,4	30,0	31,8	17,7	18,9	29,6	32,0
4	44,0	49,1	25,0	28,6	34,8	38,9	20,2	23,2	32,0	37,7
5	47,7	55,8	27,5	33,5	38,6	45,2	22,2	27,3	33,8	42,8
6	50,6	61,7	29,6	38,0	41,8	50,9	23,9	31,2	35,3	47,6
p	0,221		0,118		0,151		0,096		0,192	
d	0,534		0,492		0,417		0,500		0,651	
Erreur type	0,0741		0,0551		0,0603		0,0475		0,0823	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Parutions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
3	49,4	52,1	32,6	34,6	41,2	43,5	32,2	34,4	52,5	55,3
4	54,1	60,0	36,5	41,1	46,0	51,4	35,6	40,9	56,6	62,6
5	57,7	66,7	39,6	46,9	49,7	58,1	38,3	46,7	59,7	68,7
6	60,5	72,2	42,1	52,2	52,7	64,0	40,4	51,9	62,2	73,9
p	0,312		0,194		0,238		0,194		0,360	
d	0,635		0,588		0,555		0,593		0,708	
Erreur Type	0,0806		0,0673		0,0763		0,0765		0,0809	

Le Modèle d'Accumulation Binomial Markovien

Estimation de la Distribution des Occasions de Voir

Enquête 1953	LIFE		POST		LOOK		L.H.J.		WEEK	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	49,4	38,3	70,4	62,0	58,2	49,1	76,1	68,8	64,7	58,4
1	17,4	24,2	12,6	12,6	19,2	26,1	10,2	15,9	9,0	16,7
2	12,0	17,5	6,8	10,4	9,8	14,5	4,5	8,4	4,5	12,8
3	8,0	10,8	3,0	5,1	5,0	6,7	4,0	4,1	6,0	8,3
4	3,0	5,7	1,5	2,3	3,0	2,6	1,5	1,8	6,0	5,2
5	4,2	2,5	1,5	0,9	2,2	0,8	1,8	0,7	3,6	3,0
6	6,0	1,0	4,2	0,4	2,6	0,2	1,9	0,3	6,2	2,2
Chi Deux	16,0		11,8		9,9		9,3		25,5	

Enquête 1958	LIFE		POST		LOOK		L.H.J.		DIGEST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	39,5	27,8	57,9	47,8	47,3	36,0	59,6	48,1	37,8	26,1
1	16,7	20,7	15,1	20,2	17,8	23,7	12,4	19,9	15,0	17,4
2	11,0	18,5	8,2	14,1	11,3	17,9	8,6	14,0	9,5	16,7
3	9,4	19,9	4,9	8,9	7,3	11,6	4,9	8,9	6,9	14,4
4	6,7	9,7	5,2	5,1	6,4	6,5	5,2	5,1	8,5	11,2
5	6,6	5,6	2,5	2,5	4,4	3,0	4,0	2,6	6,8	7,7
6	10,1	3,3	6,2	1,4	5,5	1,3	5,3	1,4	15,5	6,5
Chi Deux	18,4		14,8		14,8		16,7		23,9	

Le modèle surestime gravement la couverture nette réelle pour les dix magazines. L'erreur type de prédiction est en moyenne 0,07. C'est beaucoup mieux que pour le modèle binomial simple mais c'est bien pire que pour la plupart des autres modèles stochastiques. Il apparaît donc qu'il existe davantage de corrélation entre les audiences des parutions successives que ne peut en expliquer le modèle Binomial Markovien.

La distribution estimée des occasions de voir n'est pas satisfaisante, sa forme continuellement décroissante contraste avec la forme en U de la distribution réelle. Le Chi Deux moyen est de 16,1. C'est une valeur supérieure à celle obtenue par tous les autres modèles stochastiques, exception faite pour le modèle de Morgensztern et le modèle Binomial simple. Aucune des distributions estimées n'obtient un Chi Deux inférieur à la valeur seuil de 2,2 correspondant au seuil de confiance de 90 %.

En résumé, il apparaît que la suppression de l'hypothèse de stationnarité d'une parution à une autre n'améliore pas la qualité de l'ajustement autant que ne le fait la suppression de l'hypothèse d'homogénéité de la population. Ce résultat n'implique pas toutefois qu'iln'existe pas un certain degré de variabilité, dans les probabilités individuelles, d'une parution à une autre. Il implique seulement que cet effet, si il existe, a moins d'importance dans la détermination de la distribution des occasions de voir que l'hétérogénéité des probabilités d'exposition d'un individu à un autre. D'un point de vue pragmatique, quand les données disponibles ne permettent pas d'estimer plus de deux paramètres, on a intérêt à choisir un modèle qui tient compte de l'hétérogénéité des individus, comme le modèle Bêta Binomial, plutôt qu'un modèle qui tient compte de l'exposition à la parution précédente.

Néanmoins, il est théoriquement possible de tester l'hypothèse de non stationnarité des probabilités individuelles d'exposition, indépendemment du facteur d'hétérogénéité de la population. Le test approprié est décrit ci-dessous, mais nous n'avons pas pu disposer des données d'un panel de lecteurs permettant de l'exécuter.

Supposons que les membres d'un panel de lecture aient enregistré leur exposition à quatre parutions d'un support donné. Nous pouvons classer ces individus selon leur séquence d'exposition aux trois premières parutions, ainsi que le montre la table suivante.

Nombre d'expositions	Séquence d'exposition	Nombre d'individus	Individus exposés à la 4e Parution	Probabilité Conditionnelle d'exposition à la 4e Parution
3	1 1 1	N 1 1 1	N(1/1 1 1)	P(1/3) = N(1/111)/111
	0 1 1	N 0 1 1	N(1/0 1 1)	P(1/011)= N(1/011)/N011
2	1 0 1	N 1 0 1	N(1/1 0 1)	P(1/101)= N(1/101)/N101
	1 1 0	N 1 1 0	N(1/1 1 0)	P(1/110)= N(1/110)/N110
	1 0 0	N 1 0 0	N(1/1 0 0)	P(1/100)= N(1/100)/N100
1	0 1 0	N 0 1 0	N(1/0 1 0)	P(1/010)= N(1/010)/N010
	0 0 1	N 0 0 1	N(1/0 0 1)	P(1/001)= N(1/001)/N001
0	0 0 0	N 0 0 0	N(1/0 0 0)	P(1/000)= N(1/000)/N000

On compte le nombre d'individus appartenant à chacune des séquences et parmi eux, on dénombre les individus qui ont aussi été exposés à la quatrième parution. Avec ces données, on peut estimer la probabilité empirique d'exposition à la quatrième parution pour chacune des séquences séparément. Si les probabilités individuelles d'exposition sont réellement stationnaires, toutes les séquences comprenant le même nombre d'expositions doivent donner des probabilités identiques d'exposition à la quatrième parution. Ainsi l'hypothèse nulle du test est la suivante :

$$H_0 : \begin{cases} P(1/011) = P(1/101) = P(1/110) \\ P(1/100) = P(1/010) = P(1/001) \end{cases}$$

L'hypothèse alternative prévoit que l'exposition à la parution précédente accroît la probabilité d'exposition à la parution suivante.

$$H_a : \begin{cases} P(1/011) = P(1/101) > P(1/110) \\ P(1/100) = P(1/010) < P(1/001) \end{cases}$$

Pour exécuter le test, nous devons calculer les probabilités théoriques d'exposition à la 4ième parution sous l'hypothèse H_0

$$\begin{cases} \hat{P}_2 = P(1/2 \text{ sur } 3) = \{N(1/011)+N(1/101)+(1/110)\}/\{N011+N101+N110\} \\ \hat{P}_1 = P(1/1 \text{ sur } 3) = \{N(1/100)+(1/010)+N(1/001)\}/\{N100+N010+N001\} \end{cases}$$

En utilisant ces deux probabilités, nous pouvons calculer le nombre attendu d'individus pour les 6 séquences sous l'hypothèse nulle, et former ensuite le test du Chi Deux suivant :

$$\text{Chi Deux} = \frac{(N(1/011) - \hat{P}_2 N011)^2}{\hat{P}_2\ N011} + \ldots\ldots\ldots + \frac{(N(1/001) - \hat{P}_1 N001)^2}{\hat{P}_1\ N001}$$

Puisque les deux probabilités $\hat{P}_1$ et $\hat{P}_2$ ont été estimées à partir des données, il nous reste seulement quatre degrés de liberté et l'hypothèse H_0 sera acceptée ou rejetée, au seuil de confiance adopté, selon la valeur obtenue pour la statistique du Chi Deux.

Cependant, les données dont nous disposons ne nous permettent pas de reconstituer les séquences d'exposition, individu par individu. La question de la non stationnarité des probabilités individuelles d'exposition ne peut donc pas être tranchée de façon définitive.

Bien entendu, le modèle Binomial Markovien pourrait être modifié afin d'introduire l'hétérogénéité de la population, soit par segmentation, soit par le recours à une distribution des probabilités individuelles d'exposition. Néanmoins, dans ce cas, de nombreux paramètres doivent être estimés et il est douteux que les résultats obtenus soient supérieurs à ceux obtenus par les meilleurs modèles stochastiques d'ordre zéro.

Cette direction de recherche n'a pas été poursuivie, principalement parce qu'elle conduit à des expressions trop complexes de la distribution des occasions de voir. En fait, les difficultés des modèles de Markov composé sont telles qu'il est impossible d'obtenir des expressions directes pour la distribution des occasions de voir.

Nous avons terminé notre étude des modèles stochastiques d'accumulation d'audience. Nous allons maintenant analyser l'apport des techniques de simulation à l'estimation de la couverture nette et de la distribution des occasions de voir.

33. Les Modèles de Simulation

Tout comme les modèles stochastiques, les modèles de simulation sont bâtis sur les probabilités individuelles d'exposition à un support donné. La principale différence entre les deux approches tient au fait que les modèles de simulation doivent attribuer matériellement une probabilité d'exposition à chacun des individus fictifs simulant la population cible réelle, tandis que les modèles stochastiques se contentent d'estimer les paramètres d'une formule mathématique reproduisant au mieux la distribution observée des occasions de voir.

Ainsi l'approche par simulation est beaucoup plus ambitieuse, et par là même beaucoup plus prometteuse que l'approche stochastique. Néanmoins, comme nous le verrons bientôt, il est beaucoup plus difficile d'attribuer des probabilités d'exposition à 5000 individus fictifs que d'estimer les deux paramètres de la distribution Bêta. D'une certaine façon, les modèles de simulation comprennent davantage de paramètres qu'il n'y a d'observations disponibles. Pour s'en convaincre, il suffit de songer à la masse de données nécessaires à la construction d'une réplique fidèle du comportement d'exposition de la population américaine face à l'ensemble des supports publicitaires disponibles!

Supposons un instant que nous disposions d'un panel de 5000 individus dont la répartition géographique, par occupations, classe d'âge, statut familial, niveau économique et éducatif ... etc, donne une image fidèle du pays. Supposons de plus que, pendant 12 mois consécutifs, chacun de ces individus note journellement, sans erreurs ni omissions, l'ensemble des programmes de télévision suivis, des journaux lus, des magazines parcourus, des stations de radio écoutées et des affichages rencontrés. Supposons enfin que ces données soient mises sur ordinateur et que nous calculions pour chaque individu sa fréquence moyenne d'exposition à chacun des supports. Ceci fait, il nous resterait néanmoins à admettre que la fréquence moyenne d'exposition est égale à la probabilité d'exposition, un saut logique apparemment anodin mais qui a des conséquences importantes. Par exemple, pour un magazine mensuel, cela conduit à remplacer 12 observations par leur moyenne sans tenir compte de leur variance.

Mais en fait, personne ne dispose d'un tel panel puisque T.G.I. se contente d'une mesure ponctuelle et Simmons de deux mesures ponctuelles. Même le panel de la Société Française SECODIP, qui s'approche le plus de cette situation idéale, ne mesure les comportements d'exposition que pour une période

de 12 semaines chaque année.

Il est donc clair que le constructeur d'un modèle de simulation doit poser un certain nombre d'hypothèses sur la manière d'attribuer des probabilités d'exposition à chaque individu fictif. Quoique ces hypothèses n'aient pas toujours été clairement perçues et explicitées par les constructeurs des modèles de simulation, elles existent néanmoins et elles conditionnent la qualité des prévisions du modèle.

Les modèles de simulation peuvent être classés selon deux dimensions :

- Quelles sont les règles d'affectation des probabilités d'exposition ?
- Comment la simulation est-elle exécutée ?

Le premier problème est le plus critique. Il s'agit de savoir quelles probabilités sont affectées à qui. De toute évidence cette attribution ne peut pas être purement aléatoire. Si nous connaissons par exemple l'audience cumulée de deux parutions, nous devons rechercher la méthode d'affectation des probabilités qui reproduira le mieux ces données. Cependant, on souhaite aussi, en général, utiliser les données fournies par les questions d'habitude de lecture ou d'écoute. Dans ce cas, il faut trouver des méthodes d'affectation qui soient compatibles d'une part avec les habitudes de lecture déclarées et d'autre part avec l'audience moyenne observée d'une parution. De plus, il est souhaitable que lorsque deux méthodes d'affectation des probabilités s'ajustent avec un même succès aux données disponibles, elles produisent aussi des extrapolations identiques quel que soit le nombre de parutions considérées. Ces deux critères premettant de juger de la qualité d'une technique d'attribution des probabilités ont été habilement résumés par Schreiber (30) :

"1- Les probabilités affectées ..... doivent permettre de reproduire les données qui ont servi à les calculer.

2- La méthode d'affectation des probabilités doit conduire à une extrapolation unique quel que soit le nombre de parutions considérées"

A ces deux critères de reproduction des données et d'unicité des extrapolations, nous ajouterons une troisième exigence. La méthode d'affectation des probabilités ne doit jamais conduire à des résultats absurdes et illogiques tel qu'un nombre négatif d'individus affectés à un certain niveau de probabilité.

Le second problème traite du choix de la technique de simulation. Faut-il utiliser la technique de Monte Carlo ou les expansions binomiales ? La réponse dépend beaucoup de la méthode d'affectation des probabilités qui est utilisée. Si la méthode d'affectation donne à chaque individu fictif une probabilité différente, alors il convient d'utiliser la méthode de Monte Carlo. Si la méthode d'affectation donne la même probabilité d'exposition à tout un groupe d'individus, alors l'expansion binomiale sera plus rapide et plus économique.

331. L'Affectation Actuarielle des Probabilités d'Exposition

Cette méthode consiste à supposer que la fréquence déclarée d'exposition, mesurée par la question d'habitude de lecture, est la probabilité d'exposition de l'individu. Cette méthode simple et directe attribuera, par exemple, la probabilité d'exposition 0,5 à un individu qui déclare lire deux parutions sur quatre d'un support donné. Cette méthode a été utilisée par Gensch (50) (51) et, plus récemment, par Boyd, Claycamp et Mc Clelland (52).

Cependant, cette méthode est totalement inadéquate car, même en l'absence de toute distorsion mémorielle, un individu possédant une probabilité d'exposition de 0,5 a tout de même une chance sur 16 de ne pas être exposé à quatre parutions. De même, un individu avec une probabilité d'exposition de 0,25 a une chance sur cinq d'être exposé à deux parutions sur quatre. De manière générale, la probabilité qu'un individu, ayant une probabilité réelle d'exposition "p", soit exposé exactement K fois sur N est donnée par l'expression binomiale :

$$P(K/N,p) = \binom{N}{K} p^K (1-p)^{N-K}$$

En d'autres termes, il est bien connu que, quoique l'espérance du nombre de piles soit égale à celle du nombre de faces, on n'obtient pas systématiquement deux faces tous les quatre lancés de la pièce. La table suivante donne la probabilité qu'un individu soit exposé 0,1,2, ou 4 fois à 4 parutions, lorsque sa probabilité réelle d'exposition est 0, 1/4, 1/2, 3/4 ou 1.

Nombre d'Expositions	Vraisemblance de ce nombre d'expositions lorsque la probabilité d'exposition est :				
à 4 parutions	0	1/4	1/2	3/4	1
0	1	0,316	0,0625	0,004	0
1	0	0,422	0,250	0,047	0
2	0	0,211	0,375	0,211	0
3	0	0,047	0,250	0,422	0
4	0	0,004	0,0625	0,316	1

Il est possible d'aller plus loin et de démonter que l'affectation actuarielle des probabilités ne reproduit même pas les données d'exposition qui ont servi à les calculer, ce qui prouve que la méthode échoue sur le critère N°1 (30).

Supposons qu'ayant observé un échantillon de 1280 individus, nous enregistrions les résultats suivants :

Nombre de Parutions	0	1	2	3	4
Nombre d'individus exposés	354	184	204	184	354

Selon la méthode actuarielle d'affectation, nous conclurions que 354 individus ont une probabilité d'exposition nulle, 184 une probabilité de 0,25 ... et ainsi de suite. En appliquant l'expansion binomiale à chaque groupe séparément, nous pouvons simuler la distribution des occasions de voir pour quatre parutions. Ceci est fait dans la table suivante.

Affectation Actuarielle des Probabilités

Nombre d'expositions à 4 parutions	Probabilités d'exposition selon la méthode actuarielle 0	1/4	1/2	3/4	1	Fréquence d'exposition simulée	Pourcent. d'erreur
0	354	58,22	12,75	0,72	0	425,69	+20,2
1	0	77,63	51,00	8,62	0	137,25	-25,4
2	0	38,81	76,50	38,81	0	154,12	-24,5
3	0	8,62	51,0	77,63	0	137,25	-25,4
4	0	0,72	12,75	58,22	354	425,69	+20,2
Nombre total d'individus	354	184	204	184	354	1280	

L'avant dernière colonne donne la fréquence d'exposition simulée et la dernière colonne enregistre les écarts avec la distribution observée dont nous sommes partis pour calculer les probabilités actuarielles. Il est maintenant clairement évident que l'affectation actuarielle conduit à des résultats fortement biaisés. En fait, les fréquences d'exposition simulées surestiment les deux extrémités de la distribution des occasions de voir et sous-estiment les fréquences intermédiaires.

Il est plaisant de noter que l'un des adversaires les plus acharné des modèles stochastiques, D.H. Gensch, ne mentionne nulle part dans ses ouvrages précités, cet échec dramatique de la méthode d'affectation actuarielle, qui la rend beaucoup plus inexacte que la plupart des modèles de type Ad Hoc ou stochastique. Il est important de remarquer que le biais inhérent à la méthode actuarielle d'affectation des probabilités ne peut pas être réduit en augmentant la taille de l'échantillon d'individus interrogés. Le biais ne peut être réduit qu'en augmentant le nombre de parutions considérées dans la question d'habitude de lecture. Il faudrait au moins tripler le nombre de parutions considérées pour réduire le biais à un niveau acceptable, ce qui entraînerait très certainement une importante distorsion mémorielle anihilant toute amélioration due à l'allongement de la période d'observation.

La table suivante, extraite de l'étude précitée de Simmons (29), montre l'importance de l'erreur introduite par l'affectation actuarielle des probabilités, dans le cas de trois magazines dont la distribution des occasions de voir a été mesurée par la méthode de la lecture au numéro, en 1964.

L'Affectation Actuarielle des Probabilités

Comparaison de la couverture nette observée avec la couverture nette simulée

(Milliers d'adultes masculins)

Enquête 1964	LIFE		LOOK		POST	
Parutions	observée	simulée	observée	simulée	observée	simulée
1	16 108	16 108	15 115	15 115	11 393	11 393
2	20 165	19 151	18 891	17 946	14 068	13 401
3	22 459	20 783	21 011	19 457	15 586	14 479
4	23 981	21 758	22 360	20 353	16 525	15 117

Comparaison de la distribution des occasions de voir observée avec la distribution des occasions de voir simulée

Enquête 1964	LIFE		LOOK		POST	
Expositions	observée	simulée	observée	simulée	observée	simulée
0	33 546	35 769	35 167	37 174	33 546	35 769
1	6 087	3 900	5 394	3 584	6 087	3 900
2	4 544	3 863	4 543	3 623	4 544	3 863
3	4 149	3 172	3 710	2 954	4 149	3 172
4	9 201	10 823	8 713	10 192	9 201	10 823

Puisque l'assimilation directe de la fréquence moyenne d'exposition à la probabilité moyenne d'exposition échoue, il faut imaginer d'autres méthodes d'affectation des probabilités aux individus simulés. Il en existe

quatre :

- Premièrement, on peut admettre que, dans le cas de quatre parutions, les probabilités d'exposition 0, 1/4, 1/2, 3/4, 1, sont exactes, et on cherche ensuite à déterminer le nombre d'individus fictifs qu'il faudra affecter à chacune de ces cinq probabilités pour reconstituer exactement les données observées. C'est la méthode des probabilités fixes qui a été suggérée par Schreiber (30).

- Deuxièmement, on peut considérer que les probabilités d'expositions, ainsi que le nombre d'individus qui les possèdent, sont tous deux inconnus. C'est la méthode des probabilités empiriques qui a été présentée par Schreiber (30), Friedman (53) et Greene (24).

- Troisièmement, plutôt que d'utiliser des probabilités discrètes, nous pouvons ajuster aux données une distribution continue de probabilités et attribuer à chaque individu une probabilité d'exposition basée sur cette distribution, en fonction du nombre de parutions auxquelles cet individu déclare avoir été exposé. La distribution la plus appropriée est la distribution Bêta Binomiale. Cette méthode a été proposée par Greene (24) qui l'intitule méthode des probabilités personnelles.

- Quatrièmement, si l'on dispose à la fois des réponses à une question d'habitude de lecture et à une question de lecture de la dernière période, on peut considérer que, pour chaque classe de fréquence de lecture, la proportion d'individus exposés durant la dernière période constitue la probabilité d'exposition des individus appartenant à cette classe. C'est la méthode utilisée par T.G.I. et nous l'intitulerons méthode d'affectation à la dernière parution.

Afin d'analyser le pour et le contre de ces quatre méthodes, nous allons recourir à des données artificielles, fabriquées "sur mesure", pour lesquelles nous connaissons à l'avance et par construction les probabilités d'exposition de chacun des individus. L'avantage des données artificielles par rapport à des données empiriques réside justement dans le fait que nous connaissons parfaitement, et à l'avance, les probabilités individuelles d'exposition. Nous avons donc un critère pour juger des probabilités affectées par chacune des quatre méthodes. Par contre, si nous avions choisi d'utiliser des données empiriques, quels que fussent nos efforts, il aurait été impossible de connaître avec certitude les véritables probabilités d'exposition.

De plus, nous serons aussi en mesure d'évaluer les conséquences d'un changement dans le nombre de classes de fréquence et dans l'hétérogénéité des individus à l'intérieur de chacune des classes. Nous évaluerons également l'effet des erreurs d'interprétation dans les échelles verbales de fréquence et nous étudierons les conséquences de l'instabilité des probabilités individuelles d'exposition.

332. L'Affectation fixe des Probabilités

Cette méthode repose sur l'hypothèse que les seules probabilités d'exposition possibles sont celles données par la proportion de parutions lues parmi N. Par exemple, si nous interrogeons des individus sur leur fréquence de lecture des quatre dernières parutions, nous admettons que les seules probabilités d'exposition possibles sont 0, 1/4, 2/4, 3/4 et 4/4. En d'autres termes les probabilités d'exposition sont discrètes et au nombre de N+1 :

$$P_i = i/N \qquad \text{avec} \qquad i = 0, 1, \ldots, N.$$

Cependant, nous nous éloignons de la méthode actuarielle en admettant que le nombre d'individus possédant chacune des cinq probabilités précédentes, puisse être différent du nombre d'individus déclarant avoir été exposé à K parutions sur N. Soit E_K^N, le nombre d'individus déclarant avoir été exposés à K parutions sur N et soit $F(P_i)$ le nombre inconnu d'individus possédant réellement la probabilité d'exposition P_i. Si nous connaissions $F(P_i)$, nous pourrions calculer E_K^N par simple application de la loi binomiale :

$$(111) \quad E_K^N = \Sigma_{i=0}^N \binom{N}{K} (P_i)^K (1-P_i)^{N-K} F(P_i)$$

$$\text{avec } P_i = i/N \quad ; \quad i = 0, 1, 2, \ldots, N$$

En utilisant les valeurs observées des E_K^N, nous pouvons résoudre les N+1 équations simultanées de l'expression (111) pour les valeurs inconnues de $F(P_i)$. On peut remarquer que ce système d'équations admet une solution unique, parce que nous admettons par hypothèse que les seules probabilités d'exposition possibles sont données par $P_i = i/N$. Malheureusement, cette hypothèse n'est pas réaliste et elle conduit à des résultats absurdes sous la forme de nombres négatifs d'individus. En effet, il n'existe aucune raison, empirique ou théorique, interdisant qu'un individu possède une probabilité d'exposition de 0,27.

Contraindre cette probabilité à être égale à 0,25 est un artifice difficilement justifiable. L'exemple numérique suivant, illustrant la méthode des probabilités fixes et ses inconvénients, est tiré de l'étude précitée de Schreiber (30).

Habitude de Lecture sur 4 Parutions	0	1	2	3	4	Total
Nombre d'individus E_K^N =	61	20	14	12	13	120

L'expression (111) nous fournit les 5 équations suivantes :

$$\begin{cases} E_0^4 = 61 = F(P_0) + (0,75)^4F(p_1) + (0,50)^4F(p_2) + (0,25)^4F(p_3) \\ E_1^4 = 20 = 4(0,25)(0,75)^3F(p_1) + 4(0,50)^4F(p_2) + 4(0,25)^3(0,75)F(p_3) \\ E_2^4 = 14 = 4(0,25)^2(0,75)^2F(p_1) + 6(0,50)^4F(p_2) + 6(0,25)^2(0,25)^2F(P_3) \\ E_3^4 = 12 = 4(0,25)^3(0,75)F(p_1) + 4(0,50)^4F(p_2) + 4(0,25)(0,75)^3F(p_3) \\ E_4^4 = 13 = (0,25)^4F(p_1) + (0,50)^4F(p_2) + (0,75)^4F(p_3) + F(p_4) \end{cases}$$

En résolvant ce système, on obtient la solution suivante :

$F(P_0) = 46,44$ $F(p_1) = 46,22$ $F(p_2) = -2,66$ $F(p_3) = 24,89$ $F(p_4) = 5,11$

Pour les données de cet exemple numérique, la méthode d'affectation fixe des probabilités échoue. Elle prévoit en effet un nombre négatif d'individus possédant la probabilité d'exposition 0,5. Nous pouvons donc conclure que la méthode des probabilités fixes tout en remplissant les critère 1 et 2 de reproduction des données et d'unicité de la solution, ne vérifie pas le critère 3 de cohérence logique de la solution. Pour améliorer l'affectation des probabilités, il faut se débarrasser de l'hypothèse des probabilités fixes. C'est ce que réalise la méthode des probabilités empiriques.

333. L'Affectation empirique des Probabilités

Dans cette méthode, on admet que les individus qui appartiennent à la même classe d'habitude de lecture possèdent en commun une même probabilité d'exposition, par ailleurs inconnue. En conséquence, les probabilités P_i et

les effectifs $F(p_i)$ sont, à présent, tous les deux inconnus. Nous avons donc 2(N+1) inconnues mais seulement N+1 équations. Etant donné que nous avons davantage de paramètres à estimer que d'observations, il existe une infinité de solutions vérifiant l'expression (111) et telles que les p_i et les $F(p_i)$ soient non négatifs.

Friedman (53), qui préconise l'utilisation de cette méthode, semble être conscient du problème puisqu'il déclare : "choisissez une série de 5 probabilités d'exposition largement espacées, par exemple 0; 0,04; 0,35; 0,70; 0,99. Il n'y a pas de série correcte, il est simplement nécessaire de sélectionner des probabilités qui ne conduisent pas à des effectifs négatifs ...". Cependant, lorsque la sélection des probabilités d'exposition ne résulte que d'une suite d'essais et d'erreurs, rien ne garantit l'unicité de la solution. Deux chercheurs différents aboutirons à deux solutions différentes, tout aussi conformes aux données, mais conduisant à des extrapolations différentes. Donc, bien que remplissant les critères 1 et 3, la méthode des probabilités empiriques échoue sur le critère 2 puisqu'elle ne garantit pas l'unicité de la solution.

Par exemple, si nous retournons à l'exemple numérique donné par Schreiber, les deux séries de probabilités suivantes reproduisent parfaitement les données, tout en produisant des estimations divergentes lorsque le nombre de parutions simulées est supérieur à quatre.

L'Affectation Empirique des Probabilités

Fréquence d'Exposition	Nbre. d'individus Exposés	Affectation N°1		Affectation N°2	
		Probabi.	Effec.	Probabi.	Effec.
0	61	0,01	51,68	0,00	3,51
1	20	0,20	14,23	0,04	62,26
2	14	0,35	30,90	0,35	24,84
3	12	0,80	7,33	0,70	21,66
4	13	0,88	15,86	0,99	7,73

Couverture Nette Observée et Couverture Nette Simulée

Parutions	Couverture Nette Observée	Couverture Nette Simulée	
		Probabilités N°1	Probabilités N°2
1	34,0	34,0	34,0
2	46,7	46,7	46,7
3	54,0	54,0	54,0
4	59,0	59,0	59,0
6		65,3	65,9
8		68,9	70,8
13		73,8	79,7

L'examen de cette table fait apparaître que l'importance des divergences, selon la série de probabilités choisie, est loin d'être négligeable, sans être toutefois supérieure aux divergences observées pour les modèles stochastiques d'accumulation d'audience. L'origine de cette divergence réside dans le choix des probabilités discrètes. Le recours à une distribution continue de probabilités va permettre de remédier à ce défaut de la méthode d'affectation empirique des probabilités.

334. L'Affectation individuelle des Probabilités

Dans cette méthode, très semblable dans l'esprit aux modèles stochastiques à populations hétérogènes, nous admettons que les probabilités individuelles d'exposition sont Bêta distribuées à travers la population. La supériorité de cette approche par rapport aux deux méthodes précédentes est double.

- D'une part, une seule distribution Bêta peut être ajustée aux données, ce qui garantit l'unicité de la solution.

- D'autre part, il n'est pas nécessaire de connaître la distribution des occasions de voir pour 4 parutions, puisque la distribution Bêta peut être ajustée à partir de l'audience cumulée de deux parutions. Bien entendu, si l'on dispose des réponses à une question d'habitude de lecture, on pourra les utiliser sans difficultés en recourant à l'une quelconque des méthodes d'estimation des paramètres précédemment décrites. Cette méthode d'affectation est donc la plus universelle puisqu'elle permet d'utiliser les données collectées par Simmons aussi bien que les données collectées par T.G.I.

La méthode se déroule en deux étapes. Dans un premier temps, une distribution Bêta est ajustée aux données. Nous avons déjà discuté en détail comment estimer les paramètres A et B et nous n'y reviendrons pas ici. La deuxième étape est ce qui distingue l'affectation individuelle des probabilités du modèle d'accumulation stochastique Bêta Binomial. Elle consiste en l'affectation, à chaque individu, d'une probabilité conditionnelle au nombre d'expositions reçues. Cette affectation est effectuée de manière à ce que, lorsque tous les individus sont regroupés ensemble, leurs probabilités reproduisent exactement la distribution Bêta estimée dans la première étape. Quoique cela puisse paraître difficile à accomplir, en fait il n'en est rien.

Posons nous tout d'abord la question suivante. Si un individu a été exposé à K parutions parmi N, quelle est la probabilité que sa probabilité d'exposition soit égale à p ? Puisque la distribution Bêta est la distribution conjuguée de la distribution binomiale, la réponse à cette question est donnée par une distribution Bêta de paramètres K et N-K.

$$(112) \qquad P(p/K,N) = \beta(p/K,N-K) = \frac{(N-1)!}{(K-1)!(N-K-1)!}\, p^{K-1}(1-p)^{N-K-1}$$

avec $0 \leq K \leq N$; entiers

De plus, la probabilité que la probabilité d'exposition de l'individu i soit comprise dans l'intervalle (a,b) peut être calculée comme suit :

$$(113) \qquad P(a < p < b/K,N) = \frac{(N-1)!}{(K-1)!(N-K-1)!} \int_a^b p^{K-1}(1-p)^{N-K-1}$$

avec $0 \leq a, b \leq 1$; réels $\qquad 0 \leq K \leq N$; entiers

Nous avons raisonné jusqu'ici comme si nous ne disposions d'aucune information au sujet du support auquel l'individu i a été exposé ! Les expressions (112) et (113) représentent tout ce que nous pouvons déduire au sujet de p, sachant que nous avons observé K expositions parmi N. Mais, en fait, nous avons des informations au sujet du support, informations qui sont résumées par la distribution Bêta qui a été ajustée aux données dans la première étape. Selon la terminologie Bayesiene, cette distribution constitue la distribution a priori de la variable aléatoire p. Nous pouvons donc, par application du théorème de Bayes, calculer la probabilité a postériori que l'individu i possède une probabilité d'exposition p, sachant qu'il a été exposé K fois sur N à un support caractérisé par les paramètres A et B :

$$(114) \quad P(p/K,N,A,B) = \beta(p/K+A,N+B-K) = \frac{\Gamma(A+B+N)}{\Gamma(K+A)\Gamma(N+B-K)} p^{K+A-1}(1-p)^{N+B-K-1}$$

avec $A,B > 0$; réels $\qquad 0 \leq K \leq N$; entiers

La moyenne de la distribution a postériori, définie par l'expression (114), est égale à l'espérance mathématique de la probabilité d'exposition de tous les individus qui, comme l'individu i, ont été exposés à K parutions parmi N. Donc, plutôt que d'affecter à cette classe d'individus la probabilité K/N comme dans la méthode actuarielle, nous leur affecterons la probabilité p :

$\hat{p}$ = Espérance Mathématique de p pour les individus exposés K fois = (K+A)/(N+A+B)

Finalement, la probabilité que la probabilité d'exposition de l'individu i soit comprise dans l'intervalle (a,b) est donné par :

$$(115) \quad P(a < p < b/K,N,A,B) = \frac{\Gamma(A+B+N)}{\Gamma(K+A)\Gamma(N+B-K)} \int_a^b p^{K+A-1}(1-p)^{N+B-K-1}$$

Les formules (114) et (115) représentent, d'un point de vue Bayesien, la meilleure affectation possible des probabilités aux individus et ce pour les

deux raisons suivantes :

a- Nos trois critères d'affectation des probabilités sont vérifiés. Premièrement la méthode reproduit la distribution observée des occasions de voir aussi bien que le permet la qualité de l'ajustement Bêta Binomial. Deuxièmement, la méthode produit une affectation unique. Troisièmement, la méthode ne conduit jamais à des résultats absurdes ou illogiques tels que des probabilités ou des effectifs négatifs.

b- Lorsque l'on dispose de données permettant de tester la qualité des prédictions du modèle, la méthode d'affectation individuelle des probabilités donne d'excellents résultats. Nous avons déjà eu l'occasion de démontrer la qualité de l'ajustement du modèle Bêta Binomial pour 10 magazines. Nous allons étendre cette démonstration à des données artificielles constituant un test encore plus difficile puisque nous connaîtrons à l'avance et par construction les véritables probabilités d'exposition.

Pour cela, nous créons trois populations de 1000 individus chacune, et nous affectons à chaque individu une probabilité d'exposition à un support fictif. Afin de rendre le test plus probant, nous utilisons une règle d'affectation différente pour chacune des trois populations. Pour la première population les probabilités sont affectées sur la base d'une mixture de deux distributions géométriques. Pour la deuxième population nous utilisons une mixture de deux distributions normales. Pour la troisième population nous utilisons une distribution normale de moyenne 0,5 et de variance 1. La description de chaque population figure dans la première ligne de chacun des trois tableaux présentant les résultats du test. Celui-ci se déroule en trois étapes :

<u>1ère Etape</u> : La couverture nette et la distribution des occasions de voir sont obtenues en simulant l'exposition à dix parutions. Elles figurent dans le tiers supérieur du tableau.

<u>2ème Etape</u> : En utilisant uniquement la couverture nette cumulée des deux premières parutions obtenue à l'étape précédente, nous appliquons la méthode d'affectation individuelle des probabilités aux 1000 individus de la population. Le résultat figure dans le tiers médian du tableau. Il permet de comparer les probabilités affectées aux probabilités réelles.

<u>3ème Etape</u> : En utilisant les probabilités affectées, nous simulons à nouveau l'exposition à dix parutions. Le résultat figure dans le tiers inférieur du tableau. En le comparant avec les données du tiers supérieur du tableau, on peut juger de la qualité de la méthode d'affectation individuelle des probabilités.

Cette procédure constitue le test le plus sévère que l'on puisse imaginer pour les deux raisons suivantes :

- Les probabilités individuelles d'exposition ne suivent pas , à dessein, une distribution Bêta, alors que dans la réalité nous savons qu'elles s'en approchent.

- Nous n'utilisons que la seule couverture nette des deux premières parutions pour réaliser l'affectation individuelle des probabilités et non la distribution des expositions à quatre parutions qui permettrait une affectation plus précise.

TEST N°1 : MIXTURE DE 2 DISTRIBUTIONS GEOMETRIQUES

SAMPLE TEST 1 MIXTURE OF 2 GEOMETRIC DISTRIBUTIONS

SAMPLE PROBABILITY DISTRIBUTION

0- 5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-
300	210	140	90	64	48	27	18	12	9	7	6	5	8	8	7	9	11	14	

ISSUES	REACH	COVER	FREQ.	FREQUENCY DISTRIBUTION 0	1	2	3	4	5	6	7	8	9	10
1	100	17.97	1.00	820	180									
2	276	27.62	1.30	724	193	83								
3	346	34.62	1.55	652	216	74	59							
4	406	40.61	1.77	594	232	84	42	48						
5	454	45.44	1.94	546	241	97	43	31	42					
6	495	49.55	2.13	505	246	108	49	28	26	38				
7	531	53.09	2.37	469	248	116	57	30	21	24	34			
8	562	56.14	2.56	438	248	125	64	33	21	18	22	31		
9	589	58.92	2.79	411	246	131	70	36	22	16	16	21	29	
10	613	61.35	2.93	387	243	135	76	43	25	16	14	15	20	27

PERSONAL PROBABILITY ASSIGNEMENT

PROBABILITY CLASSES	TRUE DISTRIB	RECOVERED DISTRIBUTION	PERSONAL 0	1	2	3	4
.00 TO .05	300	433	671	189	8	0	0
.05 TO .10	210	111	134	148	30	3	0
.10 TO .15	140	76	73	137	92	8	1
.15 TO .20	90	59	45	125	70	16	2
.20 TO .25	64	48	28	100	84	27	4
.25 TO .30	48	40	18	82	92	38	7
.30 TO .35	27	35	12	76	96	51	11
.35 TO .40	18	30	8	68	96	63	18
.40 TO .45	12	27	5	47	92	74	26
.45 TO .50	9	24	3	35	84	83	35
.50 TO .55	7	21	2	26	75	98	47
.55 TO .60	6	18	1	14	64	94	59
.60 TO .65	5	16	1	12	52	94	73
.65 TO .70	5	14	0	7	40	98	87
.70 TO .75	6	12	0	4	29	82	101
.75 TO .80	7	11	0	2	19	70	113
.80 TO .85	9	9	0	1	11	55	121
.85 TO .90	11	7	0	0	5	37	122
.90 TO .95	14	5	0	0	2	19	110
.95 TO 1.00	20	3	0	0	0	4	67
MEAN PROBA	.1797	.1797	.0570	.2274	.3970	.5666	.7363

BETA PARAMETERS

A =	.3406	.3406	1.3406	2.3406	3.3406	4.3406
B =	1.5549	3.5549	4.5549	3.5549	2.5549	1.5549

CHI SQUARE TEST OF TRUE VERSUS RECOVERED DISTRIBUTION 3.29

RECOVERED
SAMPLE PROBABILITY DISTRIBUTION

0- 5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-
433	111	76	59	40	49	39	30	27	24	21	18	16	14	12	11	9	7	5	

ISSUES	REACH	COVER	FREQ.	FREQUENCY DISTRIBUTION 0	1	2	3	4	5	6	7	8	9	10
1	186	18.57	1.00	814	186									
2	288	28.78	1.29	712	204	83								
3	397	39.66	1.56	643	206	108	58							
4	488	48.79	1.82	592	205	105	64	34						
5	449	44.96	2.07	551	203	106	69	45	25					
6	482	48.22	2.31	518	202	106	71	50	34	20				
7	511	51.08	2.54	489	200	105	72	53	39	27	16			
8	536	53.57	2.77	464	199	104	72	54	41	31	22	13		
9	558	55.73	3.00	442	198	103	72	54	43	33	25	18	11	
10	578	57.76	3.21	422	198	102	71	55	43	35	28	21	16	9

TEST N°2 : MIXTURE DE 2 DISTRIBUTIONS NORMALES

TEST 2 MIXTURE OF TWO NORMAL DISTRIBUTIONS

PROBABILITY DISTRIBUTION

5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100	TOTAL
162	139	103	73	46	27	15	7	3	1	2	5	9	15	25	35	45	54	60	1000

REACH	COVER	FREQ.	FREQUENCY DISTRIBUTION 0	1	2	3	4	5	6	7	8	9	10
317	31.72	1.00	683	317									
423	42.26	1.58	577	211	212								
492	49.16	1.94	508	207	110	175							
545	54.53	2.33	455	215	91	85	154						
589	58.93	2.69	411	220	97	56	78	138					
626	62.62	3.04	374	222	106	51	45	76	126				
658	65.77	3.38	342	220	116	56	34	43	74	119			
685	69.47	3.71	315	216	123	63	33	26	43	73	106		
708	70.82	4.03	292	211	120	70	36	23	27	46	71	98	
729	72.87	4.35	271	206	131	77	41	23	19	27	44	69	91

PERSONAL PROBABILITY ASSIGNEMENT

ITY SES	TRUE DISTRIB	RECOVERED DISTRIBUTION	PERSONAL PROBABILITY ASSIGNEMENT 0	1	2	3	4
.05	178	338	674	91	5	0	0
.10	162	58	124	117	19	1	0
.15	139	56	69	116	34	4	0
.20	103	45	44	112	48	7	0
.25	73	39	29	102	68	13	1
.30	46	34	20	91	78	19	1
.35	27	31	14	79	78	27	3
.40	15	29	9	68	82	36	4
.45	7	28	6	56	89	43	7
.50	3	26	4	46	84	55	10
.55	1	25	3	36	81	64	14
.60	2	25	2	28	76	76	20
.65	5	24	1	20	69	83	27
.70	9	24	1	14	60	90	37
.75	15	25	0	9	50	98	49
.80	25	25	0	6	39	97	66
.85	35	26	0	3	28	95	89
.90	45	29	0	1	18	88	122
.95	54	33	0	0	6	71	176
1.00	60	56	0	0	2	37	375
OBA	.3172	.3172	.0688	.2629	.4658	.6670	.0691

PARAMETERS

.3009	.3009	1.3009	2.3009	3.3009	4.3009
.6476	4.6476	3.6476	2.6476	1.6476	.6476

RE TEST OF TRUE VERSUS RECOVERED DISTRIBUTION 19.19

ED

PROBABILITY DISTRIBUTION

5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100	TOTAL
80	56	45	33	34	31	29	28	26	25	25	24	24	25	25	26	29	33	56	1000

REACH	COVER	FREQ.	FREQUENCY DISTRIBUTION 0	1	2	3	4	5	6	7	8	9	10
322	32.17	1.00	678	322									
432	43.22	1.43	568	221	211								
496	49.57	1.95	504	190	141	164							
540	53.96	2.38	460	176	117	118	137						
573	57.31	2.81	427	167	105	90	92	118					
600	60.00	3.22	400	162	97	80	75	80	105				
623	62.29	3.62	377	154	92	74	67	65	72	95			
642	64.20	4.01	353	153	88	69	61	58	58	66	86		
659	65.98	4.39	341	154	85	66	57	53	51	53	61	80	
674	67.42	4.77	326	152	82	63	54	50	47	47	48	57	74

TEST N°3 : DISTRIBUTION NORMALE DE MOYENNE 0,5 ET DE VARIANCE 1

SAMPLE TEST 3 NORMAL DISTRIBUTION MEAN=0.5 VARIANCE=1

SAMPLE PROBABILITY DISTRIBUTION

0- 5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-1
2	5	18	18	31	49	69	98	108	118	118	109	98	69	49	31	18	18	5	2

ISSUES	REACH	COVER	FREQ.	FREQUENCY DISTRIBUTION 0	1	2	3	4	5	6	7	8	9	10
1	500	50.00	1.00	500	500									
2	723	72.26	1.38	277	445	277								
3	834	83.39	1.60	166	334	334	166							
4	894	89.43	2.24	106	242	306	242	106						
5	929	92.92	2.69	71	175	255	255	175	71					
6	951	95.05	3.16	49	126	204	237	204	126	49				
7	964	96.41	3.63	36	95	162	207	207	162	95	36			
8	973	97.31	4.11	27	72	128	176	194	176	128	72	27		
9	979	97.93	4.60	21	55	102	147	175	175	147	102	55	21	
10	984	98.36	5.08	16	44	82	122	154	165	154	122	82	44	16

PERSONAL PROBABILITY ASSIGNEMENT

PROBABILITY CLASSES	TRUE DISTRIB	RECOVERED DISTRIBUTION	PERSONAL PROBABILITY ASSIGNEMENT 0	1	2	3	4
.00 TO .05	2	0	1	0	0	0	0
.05 TO .10	5	2	18	2	0	0	0
.10 TO .15	18	9	49	13	2	0	0
.15 TO .20	18	21	98	34	9	2	0
.20 TO .25	31	37	124	63	22	6	1
.25 TO .30	49	55	144	95	44	14	3
.30 TO .35	69	74	145	121	70	29	8
.35 TO .40	98	98	131	136	98	49	17
.40 TO .45	108	103	107	136	121	75	32
.45 TO .50	118	109	79	124	134	101	53
.50 TO .55	118	109	53	101	134	124	79
.55 TO .60	108	103	32	75	121	136	107
.60 TO .65	98	98	17	49	98	136	131
.65 TO .70	69	74	8	29	70	121	145
.70 TO .75	49	55	3	14	44	95	144
.75 TO .80	31	37	1	6	22	63	124
.80 TO .85	18	21	0	2	9	34	98
.85 TO .90	18	9	0	0	2	13	49
.90 TO .95	5	2	0	0	0	2	18
.95 TO 1.00	2	0	0	0	0	0	1

MEAN PROBA .5000 .5000 .3352 .4176 .5000 .5824 .6648

BETA PARAMETERS

A =	4.0662	4.0662	5.0662	6.0662	7.0662	8.0662
B =	4.0662	8.0662	7.0662	6.0662	5.0662	4.0662

CHI SQUARE TEST OF TRUE VERSUS RECOVERED DISTRIBUTION .13

RECOVERED
SAMPLE PROBABILITY DISTRIBUTION

0- 5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-1
0	2	9	21	37	55	74	98	103	109	109	103	98	74	55	37	21	9	2	0

ISSUES	REACH	COVER	FREQ.	FREQUENCY DISTRIBUTION 0	1	2	3	4	5	6	7	8	9	10
1	500	50.00	1.00	500	500									
2	722	72.24	1.38	278	445	278								
3	834	83.36	1.60	166	334	334	166							
4	894	89.42	2.24	106	242	304	242	106						
5	930	92.96	2.69	70	177	253	253	177	70					
6	951	95.13	3.15	49	130	204	234	204	130	49				
7	965	96.52	3.63	35	98	163	204	204	163	98	35			
8	975	97.45	4.10	25	74	130	174	191	174	130	74	25		
9	981	98.09	4.59	19	57	105	147	172	172	147	105	57	19	
10	985	98.54	5.07	15	45	85	123	152	162	152	123	85	45	15

L'analyse des résultats de nos trois simulations montre l'excellence de la méthode d'affectation individuelle des probabilités. Malgré les conditions sévères du test, les distributions initiales de probabilité sont reconstituées de façon satisfaisante. Il est quelque peu stupéfiant d'observer que, sur la seule base de la couverture nette des deux premières parutions, $R_1 = 180$ et $R_2 = 276$, nous pouvons déduire que 620 individus ont une probabilité d'exposition comprise entre 0 et 0,15; alors que le nombre réel d'individus est 650. La même constatation s'applique à la 2ème population, qui présentait, elle aussi, une distribution en forme de U. Les résultats sont encore meilleurs pour la 3ème population qui est normalement distribuée. Les trois enseignements que l'on peut tirer de cette expérience sont les suivants :

1- Il est faux d'assimiler la fréquence d'exposition avec la probabilité d'exposition, même si l'on dispose de données portant sur dix parutions. Ceci est mis en évidence par la table suivante qui compare la distribution réelle des probabilités avec la distribution que l'on obtiendrait en appliquant la méthode actuarielle d'affectation des probabilités.

Probabilités	Population N°1		Population N°2		Population N°3	
Classes	Réelle	Actuarielle	Réelle	Actuarielle	Réelle	Actuarielle
0,00 - 0,05	300	387	178	271	2	16
0,05 - 0,10	350	243	297	206	15	44
0,15 - 0,25	154	135	176	131	49	82
0,25 - 0,35	67	76	73	77	118	122
0,35 - 0,45	30	43	22	41	198	154
0,45 - 0,55	16	25	4	23	236	165
0,55 - 0,65	11	16	7	19	198	154
0,65 - P,75	11	14	24	27	118	122
0,75 - 0,85	16	15	60	44	49	82
0,85 - 0,95	25	20	99	69	15	44
0,95 - 1,00	20	27	60	91	2	16

2- La moyenne, pour chaque classe de fréquence, des probabilités individuelles est différente de la probabilité actuarielle et de plus elle varie avec la constitution de la population, tandis que les probabilités actuarielles restent fixes. Cette fixité des probabilités actuarielles, quelle que soit la population considérée, est la cause principale du caractère fallacieux de cette méthode.

Nombre d'Expositions	Probabilités Actuarielles	Probabilités Bêta		
		Population 1	Population 2	Population 3
0	0,0	0,058	0,061	0,335
1	0,25	0,227	0,263	0,418
2	0,50	0,397	0,465	0,500
3	0,75	0,567	0,667	0,582
4	1,00	0,736	0,869	0,665

Cette table indique que l'intervalle séparant les probabilités extrêmes diminue lorsque la population passe d'une distribution en U à une distribution en cloche.

3- La méthode des probabilités individuelles est extrêmement robuste. Quoiqu' elle soit basée sur la distribution Bêta , elle fonctionne tout aussi bien lorsque la population n'est pas Bêta distribuée. Il en est ainsi à cause de l'extrême flexibilité de la distribution Bêta, qui s'ajuste de façon satisfaisante même à une distribution normale.

La qualité des estimations de la méthode des probabilités individuelles peut être évaluée de deux façons. Premièrement, en calculant le test du Chi Deux entre la distribution réelle et la distribution affectée des probabilités. Les résultats de ce test apparaissent au centre des trois tables précédentes Ils indiquent que la distribution affectée ne peut pas être distinguée de la distribution réelle pour la première et la troisième population. Pour la seconde population, composée d'une mixture de deux distributions normales, la valeur du Chi Deux est de 15,19, ce qui indique, avec une probabilité de 0,25, une réelle différence entre la distribution originale et la distribution affectée. La distribution affectée par la méthode des probabilités individuelles est moins concave que la distribution originale. Toutefois, il convient de noter

que cette différence n'affecte que fort peu la couverture nette et la distribution des occasions de voir.

L'objectif des modèles de simulation étant l'estimation de la couverture nette et de la distribution des occasions de voir, la seconde manière d'apprécier les performances de la méthode consiste à comparer d'une part la couverture nette originale avec la couverture nette estimée, et d'autre part la distribution originale des occasions de voir avec la distribution estimée. La table suivante effectue cette comparaison pour six parutions.

Couverture Nette Cumulée Originale et Estimée

Parutions	Population 1		Population 2		Population 3	
	Originale	Estimée	Originale	Estimée	Originale	Estimée
3	34,82	35,66	49,16	49,57	83,39	83,36
4	40,61	40,79	54,53	53,96	89,43	89,42
5	45,44	44,86	58,93	57,31	92,92	92,96
6	49,55	48,22	62,62	60,00	95,05	95,13
Erreur type	0,0084		0,0158		0,0005	

Ces résultats montrent que l'estimation de la couverture nette par la méthode des probabilités individuelles est aussi satisfaisante que celle obtenue par le modèle Bêta Binomial. La table suivante montre qu'il en est de même pour l'estimation de la distribution des occasions de voir.

Distribution des Occasions de Voir Originale et Estimée

Expositions	Population 1		Population 2		Population 3	
	Originale	Estimée	Originale	Estimée	Originale	Estimée
0	50,5	51,8	37,4	40,0	4,9	4,9
1	24,6	20,2	22,2	16,2	12,8	13,0
2	10,8	10,6	10,6	9,7	20,4	20,4
3	4,9	7,0	5,1	8,0	23,8	23,4
4	2,8	5,0	4,5	7,5	20,4	20,4
5	2,6	3,4	7,6	8,0	12,8	13,0
6	3,8	2,0	12,6	10,6	4,9	4,9
2	4,5		6,1		0,2	

En résumé, si l'on choisit de recourir à la technique de simulation pour estimer l'accumulation des audiences, la seule méthode appropriée d'affectation des probabilités est la méthode des probabilités individuelles. Jusqu'à présent, nous avons supposé que les données d'enquête, dont nous disposions pour affecter les probabilités, étaient parfaites. Or, nous savons bien qu'il n'en est rien et qu'il s'agit d'être très prudent, spécialement vis à vis des réponses à la question portant sur les habitudes de lecture ou d'écoute. Nous allons analyser maintenant les six principales sources d'erreur affectant la mesure des fréquences d'exposition. Nous montrerons aussi comment la méthode d'affectation des probabilités selon la lecture de la dernière période permet de corriger certaines d'entre elles.

334. L'affectation des probabilités selon la lecture de la dernière période

Six catégories d'erreurs sont susceptibles d'affecter les réponses données à une question portant sur les habitudes de lecture ou d'écoute. La plus évidente concerne les défaillances mémorielles des enquêtés. La seconde tient à l'existence des lectures parallèles et des lectures répétées. La troisième que nous intitulons biais de découpage, tient à ce que les

probabilités individuelles d'exposition, qui sont continues, sont transformées en probabilités discrètes, en forçant l'enquêté à choisir parmi l'une des fréquences proposées par l'intitulé de la question. La quatrième, que nous appelons biais d'agrégation, tient à l'hétérogénéité des probabilités d'exposition des individus appartenant à la même classe de fréquence. La cinquième, à qui nous donnons le nom de biais sémantique, tient aux différences d'interprétation, d'un individu à un autre, des intitulés de chaque classe de fréquence. Ce type d'erreur affecte surtout les échelles verbales d'habitude de lecture ou d'écoute. La sixième et dernière source d'erreur est liée à l'instabilité des probabilités d'exposition d'une parution à une autre.

Les catégories d'erreur N°3 et 4 sont résolues par l'utilisation de la méthode d'affectation individuelle des probabilités et nous n'y reviendrons pas. Par contre, les erreurs ayant pour origine les défaillances mémorielles et les lectures parallèles ou répétées peuvent être diminuées par la méthode d'affectation des probabilités selon la lecture de la dernière période. Le biais mémoriel n'affecte pas tous les supports dans la même mesure. Certains supports sont plus prestigieux que d'autres et ils tendent à conférer un certain statut social à ceux qui les lisent. En conséquence, leur audience tend à être surestimée par la question d'habitude de lecture. Afin d'y remédier, on peut compléter la question d'habitude de lecture par une question portant sur la lecture de la dernière semaine. L'expérience montre que la réponse à cette deuxième question sera plus objective, l'individu ayant pu satisfaire son désir de prestige et d'affiliation dans sa réponse à la question d'habitude de lecture. Cette technique est utilisée aux Etats Unis par T.G.I. et en France par le C.E.S.P. . Pour chacune des classes de fréquence de la question d'habitude de lecture, on calcule la proportion d'individus exposés durant la dernière semaine. Cette proportion est censé mesurer la véritable probabilité d'exposition des individus appartenant à cette classe de fréquence. Il s'agit en quelque sorte d'un étalonnage des fréquences déclarées sur la base des lectures de la dernière période. Toutefois, cela ne garantit pas une estimation correcte de l'accumulation d'audience car l'effet de découpage, l'effet sémantique et l'effet d'agrégation persistent.

Nous allons illustrer cette méthode par un exemple numérique. Supposons que nous connaissions, sans biais d'aucune sorte, les lectures hebdomadaires, sur une période de 12 semaines, pour un échantillon de 1000 individus, et qu'elles soient réparties comme l'indique le tableau suivant.

Distribution des Occasions de Voir			Couverture Nette Cumulée	
Nombre d'Expositions	Nombre d'Individus	Lecture de la Dernière Semaine	Parutions	Couverture Nette
0	679	0	1	123
1	60	5	2	180
2	48	8	3	216
3	44	11	4	241
4	36	12	5	260
5	29	12	6	275
6	24	12	7	286
7	17	10	8	296
8	15	10	9	304
9	12	9	10	310
10	6	5	11	316
11	12	11	12	321
12	18	18		
Total	1 000	123 = Audience hebdomadaire		

Supposons maintenant, qu'à la fin de la période de 12 semaines, nous posions à ces 1000 individus, les deux questions suivantes :

1- Combien de numéros du magazine X lisez-vous en moyenne chaque mois ?
(cochez la case correspondant à votre fréquence de lecture)

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐

2- Au cours des 7 derniers jours, avez-vous lu un numéro du magazine X ?

OUI ☐ NON ☐

Supposons enfin que chaque individu soit doué d'une mémoire infaillible et comprenne parfaitement les sens des deux questions. Bien que nous soyons dans des conditions idéales, certains individus éprouveront des difficultés à répondre à la première question. Par exemple, un individu qui a lu un numéro sur12 ne correspond à aucune des fréquences de lecture prévues. Il pourra cocher aussi bien la première que la seconde case, selon qu'il se réfère au mois où la lecture s'est produite, ou aux deux mois où il n'a rien lu. C'est ce que nous appelons l'effet de découpage.

Pour illustrer le fonctionnement de la méthode d'affectation des probabilités selon la lecture de la dernière période, nous allons supposer que tous les individus interprètent de façon identique les cinq échelons de fréquence de lecture. Ceux qui ont lu 0 ou 1 numéro sur 12 cochent la première case, ceux qui ont lu 2,3 ou 4 numéros cochent la seconde case, et ainsi de suite comme l'indique la table suivante.

Résultats de l'Enquête

Fréquence de lecture	Interprétation Commune	Nombre d'individus	Lecture de la dernière semaine	Probabilités de lecture
0 sur 4	0,1 sur 12	739	5	5/739
1 sur 4	2,3,4 sur 12	128	31	31/128
2 sur 4	5,6,7 sur 12	70	34	34/70
3 sur 4	8,9,10 sur 12	33	24	24/33
4 sur 4	11, 12 sur 12	30	29	29/30
	Total =	1 000	123 = Audience hebdomadaire	

Comme l'indique cette table, la méthode consiste à calculer, pour chaque classe de fréquence, la probabilité de lecture de la dernière période, et à admettre que tous les individus qui se sont rangés dans cette classe possèdent cette probabilité de lecture. Cette méthode satisfait les critères 2 et 3 puisque l'affectation est unique et qu'elle ne conduit jamais à des effectifs négatifs. Hélas, la méthode ne satisfait pas au critère N°1 de reproduction des données. Pour le montrer il suffit de simuler l'exposition à 12 parutions et de comparer

la couverture nette et la distribution des occasions de voir simulées avec les données originales. La couverture nette simulée est obtenue par la formule suivante :

$$(116) \qquad R_N = \Sigma_{i=0}^{4} \left(1 - (1-p_i)^N\right) F(p_i)$$

avec p_i = probabilité de lecture pour la classe de fréquence i

$F(p_i)$= effectif de la classe de fréquence i

Le calcul de la couverture nette simulée pour 12 parutions nous donne :

$$R_{12} = \left(1-(734/739)^{12}\right)*739 + \left(1-(97/128)^{12}\right)*128 + \left(1-(36/70)^{12}\right)*70 + \left(1-(9/33)^{12}\right)*33 + \left(1-(1/30)^{12}\right)*30 = 316$$

Il apparaît donc que la couverture nette réelle, qui est de 321 individus, est sous-estimée de 5 individus par la méthode. Cette sous-estimation est la conséquence du découpage des probabilités individuelles en cinq probabilités dicrètes. Néanmoins, il convient de noter que la méthode d'affectation selon les lectures de la dernière période obtient un bien meilleur résultat que la méthode actuarielle qui, dans ce cas, prévoit une couverture nette de 257 individus, soit une sous-estimation de 64 individus.

$$R_{12} = \left(1-(3/4)^{12}\right)*128 + \left(1-(1/2)^{12}\right)*70 + \left(1-(1/4)^{12}\right)*33 + 30 = 257$$

La table suivante permet d'évaluer la supériorité de la méthode d'affectation selon les lectures de la dernière période par rapport à la méthode actuarielle.

Estimation de la Couverture Nette et de la Distribution des Occasions de Voir

Comparaison des Couvertures Nettes				Comparaison des Occasions de Voir			
Parutions	Observée	Actuarielle	Dernière Parution	Exposi.	Observée	Actuarielle	Dernière Parution
1	123	122	123	0	679	743	684
2	180	169	177	1	60	16	75
3	216	198	210	2	48	31	35
4	241	216	234	3	44	37	37
5	260	228	252	4	36	33	33
6	275	237	266	5	29	27	27
7	286	243	277	6	24	22	22
8	296	248	287	7	17	18	18
9	304	251	296	8	15	15	15
10	310	254	303	9	12	12	12
11	316	256	310	10	6	9	9
12	321	257	316	11	12	5	12
				12	18	31	21
Erreur type		0,0422	0,0064	Chi Deux		6,1	1,1

La comparaison peut aussi être faite au niveau de la reproduction des résultats des réponses à la question de fréquence de lecture, en tenant présent à l'esprit que l'interprétation sémantique des fréquences que nous avons adopté est, elle aussi, à l'origine d'une distorsion par rapport à la réalité.

Distribution des Occasions de Voir pour 4 Parutions

Parutions	Données Originales	Réponses à la question 1	Estimation Actuarielle	Estimation dernière période
0	759	739	784	766
1	102	128	73	95
2	65	70	60	60
3	39	33	38	39
4	36	30	45	40
Chi Deux		0,9	1,2	0,1

Cette dernière table mérite plusieurs commentaires :

- Tout d'abord, la distribution résultant des réponses à la question de fréquence de lecture n'est pas égale à la distribution originale. Il en est ainsi à cause de l'effet de découpage et de l'effet d'interprétation précédemment décrit. Ceci se produit, même si les individus possèdent une mémoire infaillible et comprennent parfaitement la question, car celle-ci réduit une infinité d'habitudes de lecture à cinq posibilités seulement. Toutefois, avec les données d'enquête, le phénomène passe inaperçu pour la bonne raison que l'on ne connait pas les comportements réels d'exposition. De plus, plusieurs autres sources d'erreurs affectent les réponses fournies par les enquêtés.

- La méthode d'affectation actuarielle des probabilités est incapable de reproduire les réponses à la question d'habitude de lecture et encore moins la distribution originale.

- La méthode d'affectation selon la lecture de la dernière période reproduit de façon satisfaisante la distribution originale, quoiqu'elle ne reproduise pas les réponses à la question d'habitude de lecture. C'est là son grand mérite. L'erreur type de prédiction de la couverture nette est sept fois plus faible que celle de la méthode actuarielle. Le test du Chi Deux est six fois plus faible dans le cas de douze parutions, et il est douze fois plus faible dans le cas de quatre parutions.

Nous allons montrer maintenant que la supériorité de l'affectation selon la lecture de la dernière période est encore plus évidente lorsqu'il existe une tendance systématique de la part des individus à gonfler leurs réponses à la question de fréquence de lecture. Pour cela nous supposerons, bien entendu, que le gonflement affecte uniquement les réponses à la question de fréquence de lecture et non les réponses à la question de lecture de la dernière période. L'importance du gonflement est arbitrairement définie comme suit : 97 individus parmi les 679 non exposés déclarent avoir été exposés 1 fois. Ceux qui ont été exposés de une à trois fois sur 12 parutions déclarent avoir été exposés 1 fois sur 4, et ainsi de suite comme l'indique la table suivante :

Résultats de l'Enquête en présence d'un gonflement systématique

Fréquence de Lecture Déclarée	Interprétation Gonflée	Nombre d'individus	Lecture de la dernière semaine	Probabilité de lecture
0 sur 4	0 sur 12 - 97	582	0	0
1 sur 4	1,2,3 sur 12	249	24	24 / 249
2 sur 4	4,5,6 sur 12	89	36	36 / 89
3 sur 4	7,8,9 sur 12	44	29	29 / 44
4 sur 4	10,11,12 /12	36	34	34 / 36
	Total	1 000	123 = audience hebdomadaire	

La couverture nette réelle est toujours égale à 321 mais, à cause du gonflement des réponses à la question de fréquence de lecture, la méthode actuarielle prévoit maintenant une couverture nette de 410 individus, ce qui représente une surestimation de 89 individus.

$$R_{12} = \left(1-(3/4)^{12}\right)*249 + \left(1-(1/2)^{12}\right)*89 + \left(1-(1/4)^{12}\right)*44 + 36 = 410$$

La méthode actuarielle est donc très sensible au gonflement des réponses. Par contre, la méthode de la dernière période y est peu sensible puisqu'elle prévoit dans ce cas une couverture nette de 344 individus, c'est à dire une surestimation modeste de 23 individus seulement.

$$R_{12} = (1-(225/249)^{12})*249+(1-(53/89)^{12})*89+(1-(15/44)^{12})*44+(1-(2/36)^{12})*36$$
$$= 344$$

En conséquence, si l'on admet que les réponses à la question de lecture de la dernière période sont peu affectées par la tendance à surestimer la fréquence d'exposition aux supports les plus prestigieux, la méthode d'affectation des probabilités selon les lectures de la dernière période constitue une protection efficace contre ce type d'erreur. Elle n'est pas la seule à jouir de cette propriété. En effet, on peut tout aussi bien recourir à la méthode des probabilités personnelles. Il suffit d'utiliser, dans l'estimation des paramètres Bêta, non pas les réponses à la question de fréquence de lecture, mais tout simplement la distribution des probabilités de lecture de la dernière période. Si p_K représente la probabilité de lecture de la dernière période pour la K ième fréquence d'exposition et si O_K^N représente l'effectif de la K ième fréquence d'exposition, on obtient :

$$\bar{K} = \Sigma_{K=0}^{N} p_K O_K^N$$
$$\sigma^2 = \Sigma_{K=0}^{N} K^2 O_K^N - (\Sigma_{K=0}^{N} K*O_K^N)^2$$

Les valeurs des paramètres A et B peuvent alors être estimés par la méthode des moments en fonction des valeurs de $\bar{K}$ et σ^2. En utilisant cette méthode, nous obtenons une couverture nette estimée de 324 individus, un résultat presque parfait puisqu'il représente une surestimation de 3 individus seulement! Il apparaît donc que la méthode des probabilités individuelles, appliquée aux probabilités de lecture de la dernière période, corrige à la fois l'effet de découpage et l'effet de gonflement. Une comparaison complète des méthodes actuarielles, de la dernière période et des probabilités individuelles est présentée dans la table suivante.

Estimation de la Couverture Nette et de la Distribution des Occasions de Voir en présence d'un gonflement systématique des réponses à la question d'habitude de lecture

Comparaison des Couvertures Nettes					Comparaison des Occasions de Voir				
Paru.	Obs.	Act.	Dern. Pério.	Prob. Indiv.	Exp.	Obs.	Act.	Dern. Pério.	Prob. Indiv.
1	123	176	123	123	0	679	590	656	676
2	180	253	178	177	1	60	32	96	83
3	216	301	213	210	2	48	59	61	47
4	241	333	240	234	3	44	69	32	34
5	260	356	261	253	4	36	59	24	27
6	275	372	278	268	5	29	43	23	22
7	286	384	293	281	6	24	32	21	20
8	296	393	306	292	7	17	25	18	17
9	304	399	317	301	8	15	20	15	16
10	310	404	327	310	9	12	16	11	15
11	316	407	336	317	10	6	12	10	14
12	321	410	344	324	11	12	6	15	14
					12	18	37	18	15
Erreur type		0,892	0,0114	0,0046	Chi Deux		10,1	3,8	2,8

Cette table montre clairement, tant du point de vue de la couverture nette que du point de vue de la distribution des occasions de voir, que la méthode des probabilités individuelles est supérieure à la méthode des probabilités de la dernière parution, laquelle est elle-même bien supérieure à la méthode actuarielle.

Examinons maintenant la dernière source d'erreur qui tient à l'instabilité des probabilités individuelles d'exposition d'une parution à une autre.

Ce problème a déjà été envisagé lorsque nous avons analysé le modèle Binomial Markovien. Ici nous allons étudier l'effet de l'instabilité des probabilités en recourant directement à la simulation de cette instabilité, comme l'a fait récemment Schreiber (54). Cet auteur admet que la distribution des probabilités individuelles d'exposition reste fixe d'une parution à une autre tandis qu'une proportion donnée d'individus instables échangent leurs probabilités d'exposition. Sur la base de cette hypothèse, Schreiber montre que lorsque la proportion d'individus instables s'accroît, la couverture nette s'accroît parallèlement, tandis que la distribution des occasions de voir évolue d'une forme en U vers une forme en W. Ce phénomène est décrit dans la table suivante :

Distribution des Occasions de Voir

selon la proportion d'individus instables

Nombre d'Expositions	Pourcentage d'Individus Instables					
	0 %	10 %	20 %	30 %	40 %	50 %
0	0,397	0,360	0,324	0,287	0,250	0,214
1	0,081	0,086	0,091	0,096	0,101	0,106
2	0,48	0,069	0,091	0,112	0,134	0,155
3	0,051	0,074	0,098	0,122	0,146	0,169
4	0,069	0,081	0,094	0,107	0,120	0,133
5	0,092	0,090	0,088	0,086	0,084	0,082
6	0,132	0,120	0,108	0,097	0,085	0,074
7	0,131	0,118	0,105	0,092	0,079	0,067
Couverture Nette	0,603	0,640	0,676	0,713	0,750	0,786
% d'accroissement	-	6,1%	12,1%	18,2%	24,4%	30,3%

D'après Schreiber R.J. (54)

Il convient de remarquer que l'hypothèse de la fixité de la distribution des probabilités individuelles d'exposition est indûment restrictive. D'autre part Schreiber admet aussi que les variations de probabilités se produisent purement au hasard. C'est-à-dire, si la probabilité d'un individu s'est accru à la parution précédente, cela n'implique pas qu'elle s'accroitra encore à la parution suivante. La variation de la probabilité d'exposition n'est ni progressive, ni à sens unique. Ce modèle d'instabilité pourrait s'intituler "instabilité Browniene". Toutefois, il paraît invraisemblable que la probabilité d'exposition d'un individu passe de 0,9 à 0,1 puis remonte à 0,8 pour redescendre ensuite. Nous opterons pour une hypothèse totalement opposée selon laquelle les variations de probabilité doivent être graduelles et relativement stables en ce qui concerne la direction de changement. En d'autres termes, nous admettrons que la variation de la probabilité s'explique par un processus d'apprentissage et qu'elle ne peut donc pas être totalement aléatoire. En suivant cette ligne de raisonnement, nous avons construit un modèle de variation des probabilités qui comprend 3 segments. Un segment de lecteurs stables dont les probabilités restent stationnaires d'une parution à une autre. Un segment de lecteurs nouveaux qui débutent avec une faible probabilité d'exposition laquelle s'accroit ensuite graduellement d'une parution à une autre au fur et à mesure que ces lecteurs deviennent de plus en plus fidèles au support. Un segment de lecteurs anciens qui débutent avec une forte probabilité d'exposition, laquelle diminue ensuite graduellement au fur et à mesure que ces lecteurs perdent leur intérêt pour le support.

La simulation a été exécutée selon les spécifications suivantes. Les lecteurs stables représentent Ws % de la population et leurs probabilités d'exposition suivent une distribution bêta de paramètres A=0,3 et B=0,7. L'audience d'une parution comprend donc 30% de ces lecteurs stables. Les lecteurs nouveaux représentent Wn % de la population et leurs probabilités d'exposition suivent une distribution Bêta dont la moyenne s'accroît de 0,1 à 0,9 lorsque l'on passe de la première à la neuvième parution. Les lecteurs anciens sont Wa % de la population et la moyenne de la distribution Bêta de leurs probabilités d'exposition décroît de 0,9 à 0,1.

Les résultats de cette simulation apparaissent dans la table suivante. Nous avons fait varier Ws de 0 % à 50 % et nous avons pris Wn = Wa = (1-Ws)/2.

Simulation de l'instabilité des probabilités

selon un modèle d'apprentissage

Nombre d'Expositions	Pourcentage d'individus instables					
	0 %	10 %	20 %	30 %	40 %	50 %
0	0,423	0,406	0,390	0,372	0,356	0,339
1	0,133	0,129	0,126	0,123	0,119	0,116
2	0,091	0,089	0,088	0,087	0,085	0,084
3	0,074	0,075	0,073	0,073	0,072	0,072
4	0,066	0,066	0,067	0,067	0,068	0,068
5	0,063	0,065	0,066	0,067	0,069	0,070
6	0,066	0,069	0,072	0,076	0,079	0,082
7	0,084	0,101	0,118	0,135	0,152	0,169
Couverture Nette	0,577	0,594	0,610	0,628	0,644	0,661
% d'accroissement	-	2,9 %	5,7 %	8,8 %	11,6 %	14,5 %

Deux conclusions peuvent être tirées de cette simulation.

-Premièrement, notre simulation confirme le résultat de Schreiber concernant l'accroissement de la couverture nette lorsque le pourcentage d'individus instables s'accroît. L'accroissement observé, par rapport à l'hypothèse de stationnarité des probabilités, est toutefois moindre lorsque l'on explique la variation des probabilités par un phénomène d'apprentissage que par une instabilité purement aléatoire. Lorsque la moitié de la population est instable, nous absorvons un accroissement de la couverture nette de 14,5% contre 30,3% pour le modèle de Schreiber.

-Deuxièmement, notre simulation ne confirme pas le résultat de Schreiber concernant la diminution du nombre d'individus fréquemment exposés lorsque le pourcentage d'individus instables s'accroît. Au contraire, nous observons que, lorsque les variations de probabilité sont dûes à une loyauté croissante ou à

un désintérêt croissant envers le support, le nombre d'individus fréquemment exposés s'accroît parallèlement au pourcentage d'individus instables. Par exemple, si nous nous reportons à la table précédente, nous observons que le pourcentage d'individus exposés 4 fois ou plus augmente lorsque le pourcentage d'individus instables s'accroît. Nous pouvons donc conclure que, si l'hypothèse d'une variation graduelle et à sens unique des probabilités donne une meilleure description du comportement réel des individus qu'une variation purement aléatoire ou qu'une absence de variation, alors la couverture nette et la fréquence moyenne d'exposition sont plus élevées que ne l'indique les modèles stationnaires d'accumulation d'audience.

Avec cette note optimiste, nous avons terminé notre étude des divers modèles d'accumulation d'audience. Il est temps maintenant d'évaluer globalement, les avantages et les inconvénients de chacun d'eux.

4. Evaluation Critique des Modèles d'Accumulation d'Audience

Notre étude a mis à jour un bon nombre de critères importants permettant de juger de la qualité d'un modèle d'accumulation d'audience. Ces critères peuvent être regroupés en trois catégories :

- Les critères théoriques qui ont trait à la nature et au nombre des hypothèses mises en avant par chaque modèle. Ces hypothèses doivent être explicites, vérifiables, cohérentes et parsimonieuses. Ces quatre critères permettent d'évaluer la validité a priori, la concision et la généralité du modèle.

- Les critères empiriques qui ont trait aux conditions d'application du modèle, à sa validité prédictive et à sa robustesse.

- Les critères pratiques qui ont trait à la nature des données requises par le modèle ainsi qu'à la facilité et au coût des calculs qui doivent être mis en oeuvre.

A partir de ces 3 catégories de critères, nous avons construit une grille d'évaluation qui comprend un total de 15 critères et qui va nous permettre d'évaluer en détail la qualité des 21 modèles d'accumulation d'audience étudiés dans ce chapitre. Cette grille est présentée dans la table suivante. Elle est suivie de la table d'évaluation proprement dite qui résume de façon synthétique les résultats de notre analyse.

15 Critères d'Evaluation pour les Modèles d'Accumulation d'Audience

A - Les Critères Théoriques

1 - L'exposition aux supports est-elle considérée comme un phénomène certain ou comme un phénomène aléatoire ?

2 - Les probabilités d'exposition sont-elles stationnaires d'une parution à une autre ?

3 - Les probabilités d'exposition sont-elles stationnaires d'un individu à un autre ?

4 - Les individus sont-ils considérés comme homogènes ?

5 - Le phénomène d'accumulation se produit-il au pur hasard ?

6 - Le modèle permet-il l'estimation de la distribution des occasions de voir ?

7 - Le modèle reproduit-il les données qui ont servi à estimer ses paramètres ?

8 - Les estimations produites sont-elles uniques ?

B - Les Critères Empiriques

9 - Les conditions d'application sont-elles difficiles à remplir ?

10 - Le modèle est-il robuste ?

11 - Les prédictions du modèle sont-elles soumises à un biais systématique ?

12 - L'erreur de prédiction est-elle inférieure à l'erreur d'échantillonage ?

13 - Le modèle reproduit-il fidèlement la distribution des occasions de voir ?

C - Les Critères Pratiques

14 - Le modèle nécessite-t-il des calculs volumineux ?

15 - Le modèle nécessite-t-il des données d'audience non habituellement disponibles ?

Evaluation de 21 Modèles d'Accumulation d'Audience

	Critères d'Evaluation														
	Théoriques								Empiriques					Pratiques	
Modèles Ad Hoc	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Logarithmique	C	0	0	0	N	N	0	0	N	N	SO	0	NA	N	N
Exponentiel *	C	0	0	0	N	N	0	0	N	N	SO	N	NA	N	0
Agostini	C	0	0	0	N	N	0	0	N	0	N	0	NA	N	N
Duplication fixe	C	0	0	0	N	N	0	0	N	0	SO	N	NA	N	N
Hofmans	C	0	0	0	N	N	0	0	N	0	SU	N	NA	N	0
Modèles Stochast.															
Binomial	A	0	0	0	0	0	0	0	N	0	SU	N	N	N	N
2 Segments	A	0	0	N	N	0	0	0	N	N	SO	N	N	N	N
3 Segments *	A	0	0	N	N	0	0	N	N	0	N	0	0	N	N
Politz	A	0	0	N	N	0	0	0	N	N	SU	N	N	N	N
Bêta Binomial	A	0	0	N	N	0	0	0	N	0	SO	0	0	N	N
Binomial Négatif	A	0	0	N	N	0	0	0	N	N	SU	N	N	N	N
Broadbent	A	0	0	N	N	0	0	0	N	0	SU	N	N	N	N
Morgensztern	A	0	0	N	N	0	0	0	0	0	SO	N	N	N	N
2 Seg. Bêta Bi. *	A	0	0	N	N	0	0	0	0	0	N	0	0	N	0
3 Seg. Bêta BI. *	A	0	0	N	N	0	0	0	0	0	N	0	0	N	0
Markov Binomial *	A	N	0	0	N	0	0	0	N	0	SU	N	N	0	N
Modèles de Simul.															
Actuariel	A	B	B	N	N	0	N	0	N	N	SO	N	N	0	0
Fixe	A	B	B	N	N	0	0	0	N	N	SO	N	N	0	0
Empirique	A	B	B	N	N	0	0	N	N	0	B	N	N	0	0
Individuel	A	B	B	N	N	0	0	0	N	0	SO	0	0	0	B
Dernière période	A	B	B	N	N	0	0	0	0	0	SU	0	0	0	0

* Modèles développés par l'auteur

Définition des Symboles

0	=	OUI	C	=	Certain
N	=	NON	A	=	Aléatoire
NA	=	Ne s'applique pas	SU	=	Surestination
B	=	Les deux alternatives s'appliquent	SO	=	Sous-estimation

Cette table peut être lue selon les lignes et selon les colonnes. Les colonnes décrivent de façon synthétique les propriétés de chacun des modèles d'accumulation. Les colonnes permettent de comparer les modèles entre eux pour chacune des propriétés. Par exemple, le modèle logarithmique est un modèle de type Ad Hoc pour lequel l'exposition est un phénomène certain, stationnaire d'une parution à une autre et d'un individu à un autre. La population cible est considérée comme étant homogène. L'accumulation ne se produit pas au hasard mais proportionnellement au logarithme du nombre de parutions. Le modèle ne permet pas d'estimer la distribution des occasions de voir, mais il reproduit l'audience cumulée des deux premières parutions à partir desquelles sont estimés ses deux paramètres. Les extrapolations du modèle sont uniques et le modèle s'applique quelle que soit l'audience cumulée des deux premières parutions à la condition qu'elles ne soient pas négatives, ce qui, bien entendu ne peut pas se produire. Toutefois, le modèle n'est pas robuste puisqu'il ne possède pas d'assymptote supérieure pour la couverture nette. Le modèle tend à sous-estimer systématiquement la couverture nette, mais son erreur type de prédiction est faible et elle se maintient à l'intérieur des bornes de l'erreur d'échantillonnage. Puisque le modèle n'estime pas la distribution des occasions de voir, le critère N°13 ne s'applique pas. Le modèle ne nécessite pas de calculs volumineux et ne requiert pas d'autres données que la couverture nette des deux premières parutions. Le même type de description est donné pour chacun des modèles par chacune des lignes de la table d'évaluation.

Cette table permet aussi de comparer les modèles entre eux, critère par critère. Par exemple, la lecture de colonne N°2, qui a trait au critère de stationnarité des probabilités d'exposition d'une parution à une autre, montre que tous les modèles de type Ad Hoc et tous les modèles stochastiques, à l'exception du modèle Binomial Markovien, adoptent l'hypothèse de stationnarité. Par contre, tous les modèles de simulation permettent d'introduire aussi bien des probabilités stationnaires que des probabilités variables.

Cette table d'évaluation permet de classer les modèles d'accumulation. Bien entendu, selon l'importance attribuée à chacun des critères, le classement sera modifié.

- Si l'utilisateur des modèles ne s'intéresse qu'à la couverture nette et non à la distribution des occasions de voir, alors, le meilleur modèle d'accumulation de type Ad Hoc est sans nul doute le modèle d'Agostini puisqu'il possède

l'erreur type de prévision la plus faible, qu'il n'est sujet à aucun biais systématique et qu'il ne requiert que la connaissance de la couverture nette cumulée des deux premières parutions. Le modèle d'Hofmans vient en seconde position, car quoique inférieur aux modèles logarithmique et exponentiel en terme d'erreur type de prévision, il ne conduit jamais à des estimations illogiques ou absurdes. Toutefois, son principal défaut est qu'il requiert la connaissance de l'audience cumulée des trois premières parutions qui est rarement disponible. Le plus mauvais modèle de type Ad Hoc est le modèle de duplication constante qui conduit à l'erreur type de prévision la plus élevée. La table suivante résume les performances des cinq modèles d'accumulation de type Ad Hoc et présente leur classement par ordre de mérite décroissant.

Classement des Modèles d'Accumulation de type Ad-Hoc

Modèle d'Accumulation	Erreur Type Moyenne	Biais Systématique	Robustesse	Données Requises	Rang
Agostini	0,0061	Non	Oui	2 parutions	1
Hofmans	0,0205	surestimation	Oui	3 parutions	2
Logarithmique	0,0116	sous-estima.	Non	2 parutions	3
Exponentiel	0,0161	Sous-estima.	Non	3 parutions	4
Duplica. Const.	0,0215	Sous-estima.	Oui	2 parutions	5

- Si l'utilisateur des modèles désire obtenir également une estimation de la distribution des occasions de voir et ne se contente pas de la seule couverture nette, alors, aucun des modèles précédents n'est satisfaisant. Le choix entre les modèles stochastiques et les modèles de simulation dépendra beaucoup des objectifs de l'utilisateur car, ainsi que nous l'avons montré, les meilleurs modèles des deux catégories ont des performances comparables. Si le temps de calcul sur ordinateur n'est pas pris en considération, et si l'on désire expérimenter avec divers degrés d'instabilité des probabilités, alors il convient de choisir parmi les modèles de simulation. Si, tout au contraire, on cherche à minimiser le temps de calcul, ce qui est nécessairement le cas lorsque l'on cherche à construire un plan publicitaire optimal, alors on devra choisir parmi les modèles stochastiques.

En ce qui concerne les modèles stochastiques, les meilleurs résultats, en terme d'erreur de prédiction, sont obtenus par nos deux modèles Bêta Binomial segmentés. La précision remarquable de leurs estimations a cependant un prix. Elle nécessite la connaissance de la couverture nette des trois ou quatre premières parutions ou de la distribution des réponses à une question d'habitude de lecture. En conséquence, lorsque seule la couverture nette des deux premières parutions est disponible, le meilleur modèle est sans conteste le modèle Bêta Binomial simple. Il est vrai que notre modèle à 3 segments obtient des performances comparables, mais il est moins satisfaisant sur le plan théorique. Tous les autres modèles stochastiques sont nettement moins satisfaisants et spécialement le modèle Binomial qui toutefois ne nécessite que la connaissance de l'audience d'une parution. La table suivante résume les performances de onze modèles stochastiques d'accumulation et présente leur classement par ordre de mérite décroissant.

Classement des Modèles Stochastiques d'Accumulation

Modèle d'accumulation	Erreur Type Moyenne	Biais Systéma.	Chi Deux Moyen	Robustesse	Données Requises	Rang
3 Segments Bêta Binomial	0,0023	Non	1,4	Oui	4 Paru.	1
2 Segments Bêta Binomial	0,0020	Non	1,8	Oui	3 Paru.	2
Bêta Binomial	0,0076	sous-esti.	1,9	Oui	2 Paru.	3
3 Segments	0,0073	Non	4,1	Oui	2 Paru.	4
Broadbent	0,0295	Surestima.	8,0	Oui	2 Paru.	5
Morgensztern	0,0530	Sous-esti.	22,2	Oui	2 Paru.	6
2 Segments	0,0570	Sous-esti.	30,3	Non	2 Paru.	7
Binomial Marko.	0,0700	Surestima.	16,1	Oui	2 Paru.	8
Binomial Négat.	0,0815	Surestima.	5,8	Non	*	9
Hétérogénéité Linéaire	0,1030	Surestima.	**	Non	2 Paru.	10
Binomial	0,1850	Surestima.	101,9	Oui	1 Paru.	11

* Ajusté à 6 parutions

** Existence de fréquences négatives

Parmi les modèles de simulation, la méthode d'affectation actuarielle des probabilités est la plus mauvaise car, d'une part, elle ne reproduit pas les données originales et, d'autre part, elle conduit à une forte sous-estimation de la couverture nette. La méthode des probabilités fixes n'est pas meilleure car elle conduit, dans certains cas, à des effectifs négatifs. La méthode des probabilités empiriques reproduit les données originales et ne conduit pas à des effectifs négatifs mais, malheureusement, elle ne garantit pas l'unicité de la solution. Finalement, seule la méthode des probabilités personnelles et l'affectation au dernier numéro sont acceptables puisqu'elles reproduisent les données originales, qu'elles ne conduisent jamais à des effectifs négatifs et qu'elles garantissent l'unicité de la solution. Cependant, la méthode des probabilités personnelles est préférable car elle permet de supprimer le biais de découpage présent dans la méthode d'affectation de la dernière période. La table suivante résume les propriétés des cinq méthodes d'affectation des probabilités et les classe par ordre de mérite décroissant.

Classement des Modèles de Simulation

Affectation des Probabilités	Reproduction des données	Unicité de la solution	Robustesse	Biais Systématique	Rang
Personnelle	OUI	OUI	OUI	NON	1
Dernière Période	OUI	OUI	OUI	Effet de découpage	2
Empirique	OUI	NON	OUI	*	3
Fixe	OUI	OUI	NON	Sous-estim.	4
Actuarielle	NON	OUI	OUI	Sous-estim.	5

* La solution n'étant pas unique, l'affectation empirique conduit aussi bien à des sous-estimations qu'à des surestimations.

Les données que nous avons utilisées jusqu'ici étaient limitées à six parutions. Cependant, trois enquêtes ont mesuré l'audience cumulée jusqu'à 12 et 13 parutions. Il s'agit des enquêtes Politz de 1950 et 1952 pour le magazine Life et de l'enquête Politz de 1956 pour le magazine Better Homes and Gardens. Ces données constituent un test plus sévère des modèles d'accumulation. Nous allons les utiliser pour montrer que notre classification des cinq meilleurs modèles stochastiques reste inchangée. Les résultats de ce test apparaissent en détail dans les huit tables suivantes.

Magazine Better Homes & Gardens - 1956 - Lecteurs âgés de plus de 10 ans

B.H.G. 1956 PEOPLE 10 OR OLDER POPULATION = 123800 ERRORS

ISSUES	OBSERVED	AGOSTINI 3 SEGMENTS	BETABI 2 SEG.BETA		AGOSTINI 3 SEGMENT	BETABI 2 SEG.BETA	
1	15500	15500	15500	15500	0	0	0
2	22550	22550	22550	22550	0	0	0
3	27250	27046	27152	27003	204	98	247
4	30800	30320	30514	30218	480	286	582
5	33550	32884	33034	32714	666	516	836
6	35800	34983	34934	34745	817	866	1055
7	37700	36755	36368	36450	945	1332	1250
8	39300	38287	37450	37915	1013	1850	1385
9	40750	39634	38267	39197	1116	2483	1553
10	42000	40834	38883	40335	1166	3117	1665
11	43100	41915	39349	41356	1185	3751	1744
12	44150	42898	39700	42281	1252	4450	1869
STANDARD ERROR =		942	2368	1319	53	(INDIVIDUALS)	

PARAMETERS

AGOSTINI EXPONENT B = .9436

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	7.41 PERCENT	.6452
OCCASIONAL	25.53 PERCENT	.2452
NON READERS	67.06 PERCENT	1.0

BETA BINOMIAL A = .1356 B = .9475

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	1.83	1.0
PROBABLE READERS	98.17	BETA WITH A = .1618 B = 1.3236

THE 3 SEGMENTS BBD CAN NOT BE FITTED

Magazine Better Homes & Gardens - 1956 - Lecteurs Masculins

B.H.G. MALE 10 YEARS OR OLDER POPULATION = 58900 ERRORS

ISSUES	OBSERVED	AGOSTINI 3 SEGMENTS	BETABI 2 SEG.BETA		AGOSTINI 3 SEGMENT	BETABI 2 SEG.BETA		
1	5550	5550	5550	5550	5550	0	0	0	0
2	8400	8400	8400	8400	8400	0	0	0	0
3	10400	10336	10363	10283	10400	64	37	117	0
4	11950	11806	11884	11674	11926	144	66	276	24
5	13150	12995	13105	12772	13152	155	45	378	-2
6	14150	13992	14095	13673	14172	158	55	477	-22
7	15000	14852	14900	14437	15043	148	100	563	-43
8	15750	15609	15554	15097	15800	141	196	653	-50
9	16400	16284	16087	15677	16469	116	313	723	-69
10	16950	16893	16520	16194	17067	57	430	756	-117
11	17400	17449	16872	16660	17607	-49	528	740	-207
12	17900	17960	17159	17084	18098	-60	741	816	-198

STANDARD ERROR = 118 342 593 109 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B = .8187

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	5.99 PERCENT	.7865
OCCASIONAL	25.26 PERCENT	.1865
NON READERS	68.75 PERCENT	1.0

BETA BINOMIAL A = .1234 B = 1.1858

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	1.33	1.0
PROBABLE READERS	98.67	BETA WITH A = .1531 B = 1.7132

THE 3 SEGMENTS BBD CAN NOT BE FITTED

Magazine Better Homes & Gardens - 1956 - Lectrices

B.H.G. FEMALE 10 YEARS OR OLDER POPULATION = 64900 ERRORS

ISSUES	OBSERVED	AGOSTINI	3 SEGMENTS	BETABI	2 SEG.BETA	AGOSTINI	3 SEGMENT	BETABI	2 SEG.BETA
1	9950	9950	9950	9950	9950	0	0	0	0
2	14150	14150	14150	14150	14150	0	0	0	0
3	16850	16732	16818	16733	16850	118	32	117	0
4	16850	18567	18701	18572	18810	283	149	278	40
5	20400	19978	20055	19986	20336	422	345	414	64
6	21650	21117	21031	21128	21576	533	619	522	74
7	22700	22068	21737	22082	22616	632	963	618	84
8	23550	22882	22246	22898	23509	668	1304	652	41
9	24350	23591	22614	23610	24289	759	1736	740	61
10	25050	24219	22880	24239	24979	831	2170	811	71
11	25700	24780	23071	24803	25598	920	2629	897	102
12	26250	25287	23210	25313	26158	963	3040	937	92

STANDARD ERROR = 666 1644 650 73 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B =1.0042

3 SEGMENTS PROPORTION EXPOSURE PROBABILITY
LOYAL READERS 8.73 PERCENT .8779
OCCASIONAL 27.59 PERCENT .2779
NON READERS 63.68 PERCENT 1.0

BETA BINOMIAL A = .1524 B = .8418

2 SEGMENTS BBD PROPORTION PROBABILITY
DEFINITE READERS 2.04 1.0
PROBABLE READERS 97.96 BETA WITH A = .1750 B = 1.1151

THE 3 SEGMENTS BBD CAN NOT BE FITTED

LIFE 1952 HOUSEHOLDS POPULATION = 42440 ERRORS

ISSUES	OBSERVED	AGOSTINI 3 SEGMENTS	BETABI 2 SEG.BETA		AGOSTINI 3 SEGMENT	BETABI 2 SEG.BETA		
1	11880	11880	11880	11880	11880	0	0	0	0
2	16080	16080	16080	16080	16080	0	0	0	0
3	18600	18470	18634	18481	18600	130	-34	119	0
4	20300	20087	20293	20112	20347	213	7	188	-47
5	21450	21287	21377	21327	21660	163	73	123	-210
6	22470	22228	22085	22262	22698	242	385	188	-228
7	23210	22996	22548	23064	23549	214	662	146	-339
8	23810	23640	22850	23721	24265	170	960	89	-455
9	24300	24192	23048	24284	24880	108	1252	16	-580
10	24720	24673	23177	24776	25416	47	1543	-56	-696
11	25080	25097	23261	25211	25890	-17	1819	-131	-810
12	25380	25476	23317	25600	26313	-96	2063	-220	-933
13	25640	25818	23353	25951	26694	-178	2287	-311	-1054

STANDARD ERROR = 159 1288 164 623 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B =1.0263

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	14.79 PERCENT	.9465
OCCASIONAL	40.40 PERCENT	.3465
NON READERS	44.81 PERCENT	1.0

BETA BINOMIAL A = .2703 B = .6945

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	4.70	1.0
PROBABLE READERS	95.30	BETA WITH A = .3141 B = .9711

THE 3 SEGMENTS BBD CAN NOT BE FITTED

Magazine Life - 1953 - Individus âgées de moins de 30 ans

LIFE 1953 UNDER 30 YEARS POPULATION = 3810 ERRORS

ISSUES	OBSERVED	AGOSTINI 3 SEGMENTS	BETABI 2 SEG.BETA		AGOSTINI 3 SEGMENT	BETABI 2 SEG.BETA		
1	1120	1120	1120	1120	1120	0	0	0	0
2	1550	1550	1550	1550	1550	0	0	0	0
3	1810	1803	1816	1798	1810	7	-6	12	0
4	1980	1977	1995	1966	1990	3	-15	14	-10
5	2110	2108	2118	2091	2124	2	-8	19	-14
6	2200	2212	2202	2168	2229	-12	-2	12	-29
7	2280	2297	2259	2267	2315	-17	21	13	-35
8	2340	2368	2298	2333	2387	-28	42	7	-47
9	2380	2430	2325	2389	2448	-50	55	-9	-68
10	2420	2483	2343	2438	2501	-63	77	-18	-81
11	2450	2531	2356	2481	2547	-81	94	-31	-97
12	2470	2573	2364	2519	2588	-103	106	-49	-118
13	2490	2611	2370	2554	2625	-121	120	-64	-135

STANDARD ERROR = 60 64 28 76 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B = .8789

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	16.04 PERCENT	.9161
OCCASIONAL	46.50 PERCENT	.3161
NON READERS	37.45 PERCENT	1.0

BETA BINOMIAL A = .3504 B = .8415

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	4.35	1.0
PROBABLE READERS	95.65	BETA WITH A = .4104 B = 1.1571

LIFE 1953 30 TO 40 YEARS POPULATION = 10310 ERRORS

ISSUES	OBSERVED	AGOSTINI	3 SEGMENTS	BETABI	2 SEG.BETA	AGOSTINI	3 SEGMENT	BETABI	2 SEG.BETA
1	2940	2940	2940	2940	2940				
2	4060	4060	4060	4060	4060	0	0	0	0
3	4730	4717	4751	4707	4730	0	0	0	0
4	5170	5170	5216	5148	5192	13	-21	23	0
5	5500	5510	5532	5474	5537	-0	-46	22	-22
6	5750	5779	5748	5731	5809	-10	-32	26	-37
7	5950	6060	5894	5939	6030	-29	2	19	-59
8	6110	6186	5994	6114	6216	-50	56	11	-80
9	6250	6346	6062	6263	6375	-76	116	-4	-106
10	6370	6486	6108	6393	6512	-96	188	-13	-125
11	6480	6609	6140	6507	6634	-116	262	-23	-142
12	6570	6720	6161	6609	6742	-129	340	-27	-154
13	6650	6819	6176	6701	6838	-150	409	-39	-172
						-169	474	-51	-188

STANDARD ERROR = 95 240 27 121 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B = .9080

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	15.51 PERCENT	.9190
OCCASIONAL	44.69 PERCENT	.3190
NON READERS	39.80 PERCENT	1.0

BETA BINOMIAL A = .3254 B = .8156

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	3.28	1.0
PROBABLE READERS	96.72	BETA WITH A = .3639 B = 1.0307

THE 3 SEGMENTS BBD CAN NOT BE FITTED

Magazine Life - 1953 - Individus de 40 à 50 ans

LIFE 1953 40 TO 50 YEARS POPULATION = 11080 ERRORS

ISSUES	OBSERVED	AGOSTINI 3 SEGMENTS	BETABI 2 SEG.BETA		AGOSTINI 3 SEGMENT	BETABI 2 SEG.BETA		
1	3300	3300	3300	3300	3300	0	0	0	0
2	4500	4500	4500	4500	4500	0	0	0	0
3	5320	5191	5233	5185	5320	129	87	135	0
4	5750	5661	5716	5648	5922	89	34	102	-172
5	6120	6011	6036	5990	6387	109	84	130	-267
6	6400	6287	6249	6258	6758	113	151	142	-358
7	6620	6512	6389	6476	7062	108	231	144	-442
8	6800	6701	6483	6659	7318	99	317	141	-518
9	6950	6863	6545	6815	7536	87	405	135	-586
10	7070	7004	6586	6950	7724	66	484	120	-654
11	7170	7129	6614	7069	7889	41	556	101	-719
12	7270	7240	6632	7176	8035	30	638	94	-765
13	7350	7340	6644	7271	8165	10	706	79	-815

STANDARD ERROR = 88 405 122 568 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B = .9493

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	15.90 PERCENT	.9364
OCCASIONAL	44.27 PERCENT	.3364
NON READERS	39.82 PERCENT	1.0

BETA BINOMIAL A = .3199 B = .7542

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	11.97	1.0
PROBABLE READERS	88.03	BETA WITH A = .6489 B = 2.5582

THE 3 SEGMENTS BBD CAN NOT BE FITTED

LIFE 1953 50 TO 60 YEARS POPULATION = 8810 ERRORS

ISSUES	OBSERVED	AGOSTINI	3 SEGMENTS	BETABI	2 SEG.BETA	AGOSTINI	3 SEGMENT	BETABI	2 SEG.BETA
1	1820	1820	1820	1820	1820	0	0	0	0
2	2510	2510	2510	2510	2510	0	0	0	0
3	2940	2914	2935	2919	2940	26	5	21	0
4	3230	3192	3221	3203	3246	38	9	27	-16
5	3450	3401	3414	3419	3481	49	36	31	-31
6	3610	3566	3546	3591	3670	44	64	19	-60
7	3740	3703	3635	3734	3827	37	105	6	-87
8	3840	3818	3696	3855	3961	22	144	-15	-121
9	3930	3917	3737	3960	4077	13	193	-30	-147
10	4000	4004	3765	4052	4179	-4	235	-52	-179
11	4060	4081	3784	4135	4270	-21	276	-75	-210
12	4110	4150	3797	4209	4352	-40	313	-99	-242
13	4160	4213	3806	4276	4427	-53	354	-116	-267

STANDARD ERROR = 35 198 57 159 (INDIVIDUALS)

PARAMETERS

AGOSTINI EXPONENT B =1.0654

3 SEGMENTS	PROPORTION	EXPOSURE PROBABILITY
LOYAL READERS	11.22 PERCENT	.9209
OCCASIONAL	32.19 PERCENT	.3209
NON READERS	56.59 PERCENT	1.0

BETA BINOMIAL A = .1890 B = .7261

2 SEGMENTS BBD	PROPORTION	PROBABILITY
DEFINITE READERS	3.45	1.0
PROBABLE READERS	96.55	BETA WITH A = .2212 B = 1.0197

THE 3 SEGMENTS BBD CAN NOT BE FITTED

Ces tables confirment les résultats obtenus pour six parutions. Dans tous les cas, notre modèle Bêta Binomial segmenté procure la meilleure estimation de la couverture nette cumulée. Son erreur moyenne de prédiction est de 223 individus contre 270 pour le modèle d'Agostini, 370 pour le modèle Bêta Binomial simple et 819 pour le modèle à 3 segments homogènes.

Nous avons ainsi achevé notre analyse du phénomène d'accumulation des audiences. Nous allons maintenant étudier, dans le prochain chapitre, le phénomène de duplication des audiences. Mais auparavant, il convient de résumer brièvement les diverses contributions que nous avons apportées à l'étude de l'accumulation des audiences.

5 . Conclusion

Dans ce chapitre nous avons démontré que, sous l'hypothèse de la constance de l'audience d'une parution à une autre, la distribution des occasions de voir et la couverture nette cumulée peuvent être converties l'une dans l'autre grâce aux formules de conversion hypergéométriques directes et inverses. Nous avons aussi démontré que la distribution des audiences fidèles peut être calculée soit à partir de la distribution des occasions de voir, soit à partir de la couverture nette cumulée, et que, inversement, ces deux dernières peuvent être calculées à partir de la distribution des audiences fidèles.

Nous avons présenté cinq modèles d'accumulation d'audience de type Ad Hoc, nous avons établi leurs hypothèses implicites et comparé leurs performances empiriques. Cette étude nous a permis d'établir un classement de ces modèles basé sur des considérations théoriques, empiriques et pratiques.

Nous avons ensuite proposé un schéma général de classification des modèles d'accumulation stochastiques basé sur la nature des hypothèses faites condernant les probabilités d'exposition. Neuf modèles d'ordre zéro ont été étudiés et comparés. Nous avons montré que l'abandon de l'hypothèse

d'homogénéité des individus permet de construire de meilleurs modèles, à la fois sur le plan théorique et sur le plan pratique. Nous avons aussi montré que les modèles à hétérogénéité individuelle conduisent à une meilleure estimation de la distribution des occasions de voir et de la couverture nette cumulée que les modèles à population segmentée. Nous avons réussi à améliorer les performances des modèles existants en introduisant un nouveau modèle à double hétérogénéité, le modèle Bêta Binomial segmenté.

Nous avons construit le premier modèle d'accumulation d'ordre un afin de tester l'hypothèse de non stationnarité des probabilités d'exposition d'une parution à une autre. Les résultats n'ayant pas été satisfaisants nous avons présenté un test exact de l'hypothèse de stationnarité.

Ensuite, nous avons analysé les modèles de simulation de l'accumulation des audiences. Nous avons discuté les mérites respectifs de cinq méthodes d'affectation des probabilités aux individus simulés sur la base des critères de reproduction des données originales, d'unicité de la solution et de robustesse des affectations. Nous avons démontré que les méthodes d'affectation actuarielle, fixe et empirique échouent sur au moins un des trois critères précédents.Les deux méthodes restantes, l'affectation personnelle et l'affectation selon la dernière période, vérifient les trois critères d'évaluation et conduisent à des estimations satisfaisantes. Enfin nous avons montré que les meilleurs modèles de simulation ne conduisent pas à des résultats supérieurs à ceux obtenus par les meilleurs modèles stochastiques.

En ce qui concerne le problème de la variabilité des probabilités d'exposition, nous avons construit un modèle d'apprentissage linéaire dont les résultats ont été comparés au modèle de variabilité aléatoire de Schreiber. Nous avons ainsi vérifié le phénomène d'accroissement de la couverture nette en fonction de la proportion d'individus instables.

Nos trois principales contributions à l'étude du phénomène d'accumulation des audiences peuvent être résumées comme suit. Nous avons conduit à son terme la première comparaison exhaustive des modèles existants, nous avons construit deux nouveaux modèles supérieurs aux modèles existants et nous avons mis au point une grille d'évaluation multicritère permettant aux utilisateurs de classer et de sélectionner les modèles de leur choix sur la base de 15 critères.

NOTES

(1) - Alfred Politz Research, Inc. Life,Study Of Consumer Expenditures, 7 volumes, Time Inc. , New York, 1958

(2) - GREENE J.D. et STOCK J.S., Advertising Reach and Frequency in Magazines, Marketmath, New York, 1967

(3) - ADAM D. "Problèmes posés par l'accumulation d'audience dans le temps", Etude IREP N.7, Paris, 1962

(4) - BROADBENT S. "Regularity of reading", Journal of Marketing Research, Aout 1964, pp. 50-58

(5) - HOFMANS P. "Measuring the cumulative Net Coverage of any combination of Media", Journal Marketing Research, Aout 1966, pp. 369-277

(6) - AGOSTINI J.M. "Analysis of Magazine accumulative audience", Journal of Advertising Research, Octobre 1962, pp. 24-27

(7) - Le modèle binomial est traité page

(8) - AGOSTINI J.M. "Le Media planning sans ordinateur", IREP Bulletin N°11 Paris, 1969

(9) - ENGELMAN F.L. "An Empirical Formula for Audience Accumulation", Journal of Advertising Research, June 1965, pp. 21-28

(10) - La courbe moyenne d'accumulation d'Engelman est construite sur la base des quatre études suivantes : enquêtes Politz de 1953 et 1958, enquêtes d'Audits & Survey de 1961 et 1962

(11) - Engelman semble être conscient de la futilité d'une approche basée uniquement sur la prise de moyennes à partir d'observations empiriques puisqu'il présente lui-même des données contradictoires provenant de l'enquête Simmons de 1964

(12) - AGOSTINI J.M. "How to estimate unduplicated audiences", Journal of Advertising Research, Vol.1, March 1961, pp. 11-14

(13) - Il faut noter qu'Agostini n'a jamais soutenu que son modèle de duplication puisse être utilisé comme modèle d'accumulation

(14) - SHERILL P.N. "The probability of advertising exposure" Journal of Advertising Research, March 1966, pp. 24-28. Voir aussi à ce sujet la discussion présentée à la section 22 de la thèse.

(15) - LEE A.M. & BURKHART A.J. "Some Optimization Problems in Advertising Media Planning", Operational Research Quaterly, vol. 11 N°3, September 1960,pp. 113-122

(16) - LEE A.M. "Decision Rules for Media Scheduling : Static Campaigns", Operational Research Quaterly, vol. 13 N°3 September 1962 pp. 229-242

(17) - TAYLOR C.J. "Some Developments in the theory and Application of Media Scheduling Methods", Operational Research Quaterly vol. 14, N°3 September 1963, pp. 291-305

(18) - ELLIS D.M. "Building Up a Sequence of Optimum Media Schedules", Operational Research Quaterly, vol. 17, N°4, December 1966, pp. 413-424

(19) - AGOSTINI J.M. "Les Questions directes sur les Habitudes de Lecture sont-elles vraiment si mauvaises", XV Esomar Conference 1962

(20) - FRANKEL L.R. "Some Recent Developments in Market Research Methods" XV Esomar Conference, 1962

- FRANKEL L.R. & SOLOMON D. "Look Audience 1962, A National Study" Audits & Surveys Co. Inc. New York 1962

(21) - HYETT G.P. "The Measurement of Readership", London School of Economics Paper, 1958, London

(22) - POLITZ A. "A study of Four Media, their Accumulative and Repeat Audiences", A. Politz Research, New York, 1953

(23) - SIMMONS W.R. "Procedures for Estimating Reach and Frequency", Annexe N°7 in 1968 Study of Selective Markets and the Media Reaching Them, W.R. Simmons & Associates Research 235 East 42th Street, New York.

(24) - GREENE " Personal Media Probabilities" Journal of advertising Research, vol.10, n°5, October 1970, pp.12-18

(25) - FRIEDMAN L. "Calculating TV Reach and Frequency", Journal of Advertising Research, vol.11 N°14, August 1971

(26) - CHANDON J.L. "Audience accumulation models : Simulation or Formulaes ?" Fifth Annual Meeting of the European Academy for Advanced Research in Marketing. Fontainebleau, April 1976

(27) - BROADBENT S. "Regularity of Reading", Journal of Marketing Research August 1964, pp. 50-58

(28) - MORGENSZTERN "Frequentation de la Presse, Généralisation, Effet Publicitaire" Etude N°19, IREP, Avril 1970, Paris

(29) - SIMMONS W.R. "An Examination of Alternative Methods of Estimating Media Reach and Frequency", W.R. Simmons & Associates Research, Inc. New York, 1968

(30) - SCHREIBER "Probability Assignements for the Simulation of Media Reach and Frequency", Journal of Advertising Research Vol.8, N°2, June 1968, pp.3-8

(31) - La notation utilisée est similaire à celle développée par Politz puis reprise par Broadbent et Greene & Stok.

(32) - CHARNES A., COOPER W.W., LEARNER D.B., SNOW E.F., "Media Mixing for distribution of frequencies with Bêta-like and lognormal distributions of frequencies", Management Science Report, N°189, Carnegie Mellan University, Janvier 1968

(33) - La distribution bêta binomiale est aussi connue sous le nom de hyperbinomial, négative hypergéométrique ou, plus simplement, binomial composée. Elle ne doit pas être confondue avec la distribution négative binomiale que nous rencontrerons plus tard.

(34) - Traduction du terme "Gross Impression"

(35) - Traduction du terme "Apparent learning"

(36) - Kendall M.G., STUART A., The Advanced Theory of Statistics, Vol.1 et 2, London, Griffin, 1969.

(37) - HEADEN R.S., "The distribution theory of net cumulative audience coverage and exposure frequency in media planning", Proceedings of the 1971 American Marketing Association Conference, pp. 230-234

(38) - ANSCOMBE F.J., "Sampling theory of the negative binomial and logarithmic distributions", Biometrica, Vol. 37, 1950, pp. 358-382

(39) - International Mathematical and Statistical Librairies

(40) - SARNDAL C.E., "Derivation of a class of frequency distribution via Boye's theorem", Journal of the Royal Statistical Society, Série B, 27, 1965, pp. 290-300

(41) - GREENE J.D., STOCK J.S., "A rate grequency model of behavior", Journal of Advertising Research, August 1971, pp.9-19

(42) - EHRENBERG AS.C, Repeat Buying Theory and Applications, North Holland, Amsterdam, 1972

(43) - FRIEDMAN L., "Calculating TV Reach and Frequency", Journal of Advertising Research, August 1971, pp. 21-25

(44) - GREENWOOD M., Yule G.V., "An enquiry into the nature of frequency distribution representative of multiple happenings", Journal of the Royal Statistical Society, Vol. 83, 1920 pp. 255-279

(45) - CHATFIELD C., "On estimating the parameters of the logarithmic series and negative binomial distributions", Biometrika, 56, 1969, pp. 411-413

(46) - HULKS B., THOMAS S.G., "PREFACE, a simple model for the prediction of television coverage and frequency distribution", XXVI ESOMAR/WAPOR Congress, Budapest 1972

(47) - CHANDON J.L., "Report on the performance of SUPERCUME, a network TV reach and frequency model", Interactive Market System Inc, New York, 1974

(48) - Nielsen Television Index - Special report on men's and women's cumulative audiences - Nielsen, 1973

CHAPITRE 6

LES MODELES DE DUPLICATION D'AUDIENCE

Le problème posé par la duplication des audiences de plusieurs supports est le suivant. Quand un plan publicitaire fait appel à plusieurs supports, avec une insertion dans chacun d'eux, combien d'individus appartenant à la population cible seront exposés au moins une fois ?

Le nombre d'individus exposés à au moins un support parmi les m supports véhiculant le message publicitaire est appelé couverture nette des m supports. Nous le noterons R_m. La connaissance de la couverture nette d'une combinaison de m supports ne suffit pas à caractériser cette combinaison. Nous désirons aussi connaître le nombre d'individus cible qui seront exposés à exactement 0,1,2,..., ou à l'ensemble des m supports. En d'autres termes, nous désirons connaître la distribution des occasions de voir générée par toute combinaison de supports.

Ce problème a suscité de nombreuses recherches. Nous avons recensé 25 articles portant sur ce sujet dans la littérature spécialisée, depuis 1960. Deux raisons peuvent expliquer cette abondance de recherches.

1- Il n'est pas possible de comparer les mérites respectifs de deux plans publicitaires si l'on ne connaît pas la couverture nette et la distribution des occasions de voir générée par ces deux plans.

2- Il existe de nombreuses données provenant des études d'audiences qui permettent de tester la validité des divers modèles de duplication par tabulation croisée des audiences de chaque support.

Dans ces conditions, pourquoi recourir à des estimations alors que des résultats exacts peuvent être obtenus par simple tabulation croisée. La

réponse tient à l'énormité du nombre des différentes combinaisons de supports(1). Cette énormité se manifeste même si l'on ne considère qu'un faible nombre de supports. Par exemple, avec 10 supports, il existe 45 paires de supports, 120 triplets de supports, 210 quadruplets, 252 quintuplets, 210 sextuplets, 120 combinaisons de 7 supports, 45 combinaisons de 8 supports, 10 combinaisons de 9 supports et une combinaison comportant les dix supports. Au total cela représente 1013 tabulations croisées. Si l'on considère que l'audience de 105 supports est actuellement mesurée par T.G.I., le nombre total de combinaisons de supports explose. Il est évidemment hors de question de tabuler 2^{105} combinaisons différentes. Même l'ordinateur le plus rapide actuellement disponible emploierait plusieurs années pour y parvenir (2). Il est donc indispensable de construire des modèles de duplication qui permettront d'estimer, avec un minimum de calcul et en recourant à un minimum de données, la couverture nette et la distribution des occasions de voir, pour n'importe quelle combinaison de supports. L'utilité d'un modèle de duplication d'audience, sa raison d'être, est d'éviter d'avoir à recourir à une coûteuse et difficile classification croisée des données d'enquête chaque fois que l'on désire connaître la couverture nette d'une combinaison de supports.

Les recherches dans ce domaine ont suivi trois directions :

a- On a tout d'abord cherché à construire des formules d'estimation Ad-Hoc, sans autre objectif que la réalisation du meilleur ajustement possible aux données disponibles. Cette direction de recherche est caractérisée par une absence d'hypothèses explicites quant aux comportements de lecture ou d'écoute des individus. Ces comportements sont supposés obéir à une loi de type déterministe. Les formules de type Ad Hoc se préoccupent exclusivement de l'estimation de la couverture nette et n'apportent aucune information sur la distribution des occasions de voir.

b- On a ensuite cherché à construire des modèles de duplication, de type stochastique, basés sur diverses hypothèses concernant les probabilités d'exposition des individus. Ces modèles peuvent être classifiés en fonction des hypothèses faites quant à l'homogénéité ou à l'hétérogénéité des individus et quant à la dépendance ou l'indépendance des audiences des divers supports. Tous les modèles stochastiques sont caractérisés par des hypothèses explicites et ils permettent tous d'estimer à la fois la couverture nette et la distribution des occasions de voir.

c- Une troisième approche est basée sur la simulation des expositions d'un échantillon d'individus auxquels sont affectés diverses probabilités

d'exposition aux différents supports. Toutefois cette approche est généralement utilisée en vue de déterminer la couverture nette et la distribution des occasions de voir de plans publicitaires comportant plusieurs insertions dans chacun des supports. A ce titre, nous l'examinerons dans le prochain chapitre.

Afin d'évaluer les divers modèles de duplication, il est nécessaire de disposer de la totalité des classifications croisées d'une enquête d'audience. Une recherche extensive n'a parmis de mettre à jour qu'une seule banque de données satisfaisant à ce critère. Il s'agit d'une étude d'audience réalisée par Marc (7). Cette étude fournit pour cinq magazines, en plus de l'audience de chacun des supports et de leur duplication par paire, une classification croisée complète des audiences. Elle nous permettra de tester les estimations des modèles de duplication pour un total de 16 combinaisons différentes de supports qui se décomposent comme suit : 10 combinaisons de 3 magazines, 5 combinaisons de 4 magazines et une combinaison de 5 magazines.

Afin de soumettre les modèles de duplication à un test plus sévère, nous utiliserons aussi une seconde banque de données portant sur un total de 11 magazines. Il s'agit de l'enquête Politz de 1964 (18). Elle nous permettra de comparer la couverture nette estimée avec la couverture nette réelle, mais nous ne pourrons pas l'utiliser pour évaluer la distribution des occasions de voir, car celle-ci n'a pas été tabulée.

1. Les Modèles Ad Hoc de Duplication

11. La Formule d'Agostini

La première formule proposée en vue d'estimer la couverture nette d'une combinaison de support est due à Jean-Michel Agostini (3), un directeur de l'agence Elvinger à Paris. Sur la base de l'enquête audience réalisée en 1957 par le CESP (4), Agostini cherche à établir une relation entre la somme des audiences de chaque support et la somme des audiences dupliquées de chaque paire de supports. Nous allons utiliser la notation suivante pour décrire sa formule :

m = Nombre de supports considérés.

A_i = Audience d'une parution du support i.

$A = \Sigma_{i=1}^{m} A_i$ = Audience brute de la combinaison des m supports.

$A_{ij} = A_i \cap A_j$ = Audience dupliquée des supports i et j

$D = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}$ = Somme des audiences dupliquées de toutes les paires de supports.

$R_m = \cup_{i=1}^{m} A_i$ = Couverture nettte de la combinaison des m supports.

Agostini étudie les variations du rapport Z de la couverture nette à l'audience brute, en fonction des valeurs du rapport X de la somme des duplications à l'audience brute.

$$Z = f(X) \quad \text{avec} \begin{cases} Z = R_m/A \\ X = D/A \end{cases}$$

Pour 98 combinaisons de magazines, extraites au hasard parmi les 32767 combinaisons contenues dans l'étude du CESP, il détermine que la fonction hyperbolique, $Z = 1/(1+1{,}125\ X)$, procure le meilleur ajustement possible. En appelant k la constante 1,125 et en remplaçant Z et X par leur définition, on obtient la formule d'Agostini :

(1) $R_m = A/(1+k\ D/A)$

En se reportant aux définitions de A et de D, on obtient :

(2) $R_m = (\Sigma_{i=1}^{m} A_i)^2 / (\Sigma_{i=1}^{m} A_i + k\ \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij})$

La constante k=1,125, baptisée constante d'Agostini, passa pour un temps pour avoir une valeur universelle. Cependant, elle ne résulte que d'un simple ajustement par la méthode des moindres carrés, opéré à partir d'un échantillon ne reflétant que les comportements de lecture propres à un certain pays et à un certain moment. Il n'y avait donc pas lieu de s'attendre à ce qu'elle fût de portée générale.

Très rapidement, les études de Kuhn (5) portant sur des données d'audience allemande, de Caffyn et Sagovsky (6) portant sur des données anglaises, de Marc (7) portant sur des revues spécialisées françaises et de Bower (8) utilisant des données canadiennes, démontrèrent la variabilité de la constante k selon les pays et selon les catégories de publications. Ces études montrèrent aussi que k, bien loin d'être une constante, varie en fonction

du nombre de supports présents dans la combinaison. En conséquence, plusieurs formules d'ajustement de k en fonction du nombre de supports furent proposées. Kuhn (5) proposa l'ajustement suivant :

$$k_m = k_{m-1} - \frac{1}{(m-(m-a)/2)^2} \qquad \text{avec } a = \begin{cases} 2 \text{ si m est pair} \\ 3 \text{ si m est impair} \end{cases}$$

Kaatz (9), utilisant les données de Nielsen de 1966, proposa de modifier comme suit la variable X = D/A :

$$X' = 1{,}006 + 0{,}053\ X$$

Ces deux formules d'ajustement ont été dérivées par des analyses de régression à partir de données d'enquête, elles n'ont donc aucun fondement théorique, tout comme la formule d'Agostini elle-même.

- Premièrement, aucune justification n'est fournie concernant la forme de la relation fonctionnelle liant Z à X, si ce n'est qu'elle permet le meilleur ajustement aux données.

- Deuxièmement, on observe que la constante k s'accroît parallèlement au nombre de supports inclus, et au rapport D/A, mais on ne fournit aucune explication quant à la cause de ce phénomène.

- Troisièmement, la formule ne permet pas d'estimer la distribution des occasions de voir.

Ces problèmes ont conduit rapidement divers chercheurs à s'interroger sur les hypothèses implicites qui se cachent derrière la formule d'Agostini. Dès 1962, Marc (10) établissait la relation liant k à la distribution des occasions de voir, puis en 1966 Hofmans (11) établissait une estimation de k en fonction des A_{ij} et enfin Calycamp et Mc Clelland (12) élucidaient la plupart des hypothèses implicites conduisant à la formule d'Agostini. Nous allons présenter ces divers résultats que nous compléterons par un nouveau résultat concernant l'hypothèse implicite faite sur les triplications en particulier et sur tous les recouvrements d'ordre supérieur à deux, en général.

Mais auparavant, il convient de rappeler un théorème important d'algèbre Booléenne qui montre que le cardinal de l'union de m évènements compatibles peut être calculé à partir des cardinaux de chacun des évènements en tenant compte des cardinaux des diverses intersections d'évènements (13).

(3) $| \cup_{i=1}^{m} A_i | = \Sigma_{i=1}^{m} |A_i| - \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} |A_{ij}| + \ldots (-1)^{m+1} \Sigma\Sigma_{i \neq j \neq m} \ldots \Sigma |A_{ij\ldots m}|$

Si les évènements considérés sont l'exposition à m supports, la formule peut être transposée comme suit :

Couverture nette = Somme des audiences - Somme des duplications + Somme des triplications - ..... $(-1)^{m+1}$ Somme des recouvrements d'ordre m.

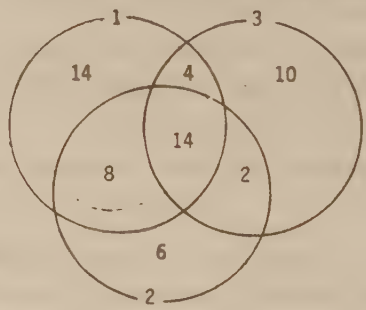
Par exemple, pour m = 2, on obtient :

(4) $R_2 = A_1 + A_2 - A_{12}$

L'exemple numérique suivant, portant sur 3 supports, nous permettra d'illustrer le théorème et de démontrer, par la même occasion, que la constante k varie, en fait, pour chaque paire de supports considérée.

Supports	Audiences
1	$A_1 = 40$
2	$A_2 = 30$
3	$A_3 = 30$
1 et 2	$A_{12} = 22$
1 et 3	$A_{13} = 18$
2 et 3	$A_{23} = 16$
1 et 2 et 3	$A_{123} = 14$

Schéma de Duplication des Audiences



Par application du théorème (3), nous obtenons.

$$R_3 = A - D + T$$

$$R_3 = (A_1 + A_2 + A_3) - (A_{12} + A_{13} + A_{23}) + A_{123}$$

$$R_3 = 100 - 56 + 14 = 58$$

Si nous considérons les supports 1 et 2, la formule d'Agostini s'écrit comme suit :

(5) $R_2 = (A_1 + A_2)^2 / (A_1 + A_2 + k\, A_{12})$

La formule exacte est donnée par l'expression (4). En combinant (5) avec (4),

nous pouvons extraire la valeur de k pour les supports 1 et 2 :

(6) $k = (A_1 + A_2) / (A_1 + A_2 - A_{12}) = 1,458$

Pour les deux autres paires de supports; nous obtenons de la même manière $k_{13} = 1,346$ et $k_{23} = 1,364$. Il est donc évident que la constante k varie selon la paire de supports et qu'elle n'est donc pas contante ! Il est aussi facile de montrer que k varie avec le nombre de supports considérés. Par exemple, avec 3 supports on obtient :

$$k = \frac{A}{D} * \frac{D-T}{A-D+T} = 1,293$$

On observe donc que k diminue avec le nombre de supports. D'une façon générale, la valeur de k pour m supports est donnée par : $k = \frac{A}{D} * \frac{A-R_m}{R_m}$

Envisageons maintenant la relation existant entre k et la distribution des occasions de voir. Nous adopterons pour cela la notation suivante :

E_i^m = Nombre d'individus exposés à exactement i supports parmi les m supports que comporte la combinaison. L'ensemble des E_i^m constitue la distribution des occasions de voir.

$S_1 = A = \Sigma_{i=1}^{m} A_i$ = Somme des audiences

$S_2 = D = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}$ = Somme des duplications

⋮

$S_r = \Sigma_{i=1}^{m-r+1} \Sigma_{j=i+1}^{m-r+2} \dots \Sigma_{r=i+r-1}^{m} A_{ij} \dots r$ = somme des recouvrements d'ordre r

Nous pouvons exprimer les E_i^m en fonction des S_r, en recourant au théorème de Waring (13) , (14).

$$E_i^m = S_i - \binom{i+1}{i} S_{i+1} + \binom{i+2}{i} S_{i+2} \dots \pm \binom{m}{i} S_m$$

(7) $$E_i^m = \Sigma_{r=i}^{m} (-1)^{i-r} \binom{r}{i} S_r$$ avec : $i = 0,1,\dots,m$

$i \leq r \leq m$, $S_0 = 1$

En extrayant de l'équation (7) la valeur des sommes S_r en fonction des E_i^m, nous obtenons :

(8) $$S_r = \Sigma_{i=r}^m \binom{i}{r} E_i^m \qquad \text{avec } r = 1,2,\ldots,m$$
$$r \leq i \leq m$$

En appliquant l'expression (8) au calcul de S_1 et de S_2, nous obtenons la somme des audiences et la somme des duplications en fonction de la distribution des occasions de voir.

(9) $$S_1 = A = E_1^m + 2\,E_2^m + 3E_3^m + \ldots. + m\,E_m^m = \Sigma_{i=1}^m\, i\,E_i^m$$

(10) $$S_2 = D = E_2^m + 3\,E_3^m + 6E_4^m + \ldots. + \binom{m}{2} E_m^m = \Sigma_{i=2}^m \binom{i}{2} E_i^m$$

Grâce à ces deux équations, nous pouvons maintenant calculer la constante k d'Agostini en fonction de la distribution des occasions de voir. Pour cela il suffit d'extraire k de la formule (1) et de remplacer A,D et R_m par leurs valeurs en fonction des E_i^m

(11) $$k = \frac{A}{D} * \frac{A-R_m}{R_m} = \frac{\Sigma_{i=1}^m\, iE_i^m}{\Sigma_{i=2}^m \binom{i}{2} E_i^m} * \frac{\Sigma_{i=2}^m\,(i-1)E_i^m}{\Sigma_{i=1}^m\, E_i^m}$$

L'expression (11) montre que k ne dépend que de la distribution des occasions de voir. Si celle-ci était connue, il serait donc possible de calculer exactement la valeur de k. Hélas, le problème ne se pose pas dans ces termes car il s'agit d'estimer k sur la seule base des duplications et des audiences qui sont les seules données disponibles. Nous présenterons dans un instant la solution préconisée par Hofmans. Auparavant nous allons montrer que, s'il est impossible de reconstituer exactement la distribution des occasions de voir en fonction des duplications et des audiences, il est du moins possible de calculer la moyenne et la variance de cette distribution.

Soit $\bar{x}$ la moyenne et σ^2 la variance de la distribution des occasions de voir. Par définition de la moyenne, nous avons :

$$\bar{x} = \Sigma_{i=o}^m\, i\,E_i^m \,/\, \Sigma_{i=o}^m\, E_i^m$$

En vertu de l'équation (9) et en notant par POP la population cible, le nombre moyen d'expositions dans l'ensemble de la population cible est donné par le

rapport de l'audience brute à la population cible :

$$(12) \quad \bar{x} = A / POP$$

Par définition de la variance, nous avons :

$$\sigma^2 = \Sigma_{i=1}^{m} i^2 E_i^m \Big/ \Sigma_{i=o}^{m} E_i^m - \bar{x}^2$$

Or, en vertu de l'équation (10), la somme des audiences dupliquées est égale à :

$$D = \Sigma_{i=2}^{m} \binom{i}{2} E_i^m$$

En notant que $\binom{i}{2} = 0$ pour $i = 1$ nous pouvons réécrire D comme suit :

$$D = \Sigma_{i=1}^{m} \binom{i}{2} E_i^m = \frac{1}{2} \left[\Sigma_{i=1}^{m} i^2 E_i^m - \Sigma_{i=1}^{m} i E_i^m \right]$$

En divisant D par la population cible nous obtenons :

$$\frac{D}{POP} = \frac{1}{2} \left[\frac{\Sigma_{i=1}^{m} i^2 E_i^m}{\Sigma_{i=o}^{m} E_i^m} - \frac{\Sigma_{i=o}^{m} i E_i^m}{\Sigma_{i=o}^{m} E_i^m} \right] = \left[\sigma^2 + \bar{x}^2 - \bar{x} \right]$$

D'où nous pouvons extraire la valeur de la variance :

$$(13) \quad \sigma^2 = \frac{2D + A}{POP} - \left(\frac{A}{POP}\right)^2$$

Les équations (12) et (13) montrent que l'on peut calculer la moyenne et la variance de la distribution des occasions de voir à partir de la seule connaissance de la somme des audiences et de la somme des duplications. Toutefois, l'annonceur est plus fréquemment intéressé par F, la fréquence moyenne d'exposition parmi les individus exposés que par x, la fréquence moyenne d'exposition dans l'ensemble de la population cible.

$$(14) \quad F = \frac{\Sigma_{i=o}^{m} i E_i^m}{\Sigma_{i=1}^{m} E_i^m} = \frac{A}{R_m}$$

En utilisant la formule (11) avec les résultats (13) et (14) nous pouvons exprimer la constante k en fonction de F et σ^2

$$(15) \qquad k = \frac{2F\,(F-1)}{F(F-1) + \sigma^2}$$

Cette formule, établie par Claycamp et Mc Clelland (12), indique que la valeur de k est fonction, à la fois, de la fréquence moyenne d'exposition F et de la variance de la distribution de voir. Cette formule nous indique aussi que, k ne peut rester constant d'une combinaison de support à une autre, que lorsque F et σ^2 vérifient la relation suivante :

$$F^2 - F = \frac{k}{2-k}\,\sigma^2$$

La formule (15) montre enfin que, lorsque la variance de la distribution des occasions de voir tend vers zéro, k tend vers 2; k est donc nécessairement compris entre zéro et deux.

Pour en finir avec les hypothèses implicites liées à la formule d'Agostini, nous allons montrer que la constance de k implique que les triplications et tous les recouvrements d'ordre supérieur à 2 peuvent être calculés à partir des audiences et des duplications.

Hofmans (11) note que la formule d'Agostini peut être développée en série géométrique infinie de base A et de raison kD/A :

$$(16) \quad R_m = \frac{A}{1 + kD/A} = A - kD + \frac{(kD)^2}{A} - \frac{(kD)^3}{A^2} + \ldots\ldots(-1)^r \frac{(kD)^r}{A^{2-1}} + \ldots.$$

Cette série ne converge que si $kD/A < 1$ c'est à dire $kD < A$ ou encore $k < A/D$. En comparant l'expression (16) avec la formule exacte (3) définissant la couverture nette, on met à jour l'hypothèse implicite concernant la valeur des recouvrements d'ordre supérieur à deux.

Indice i	Définition de la Couverture nette	Formule d'Agostini	Commentaire-
1	$S_1 = A$	A	Exact
2	$S_2 = -D$	$-kD$	Inexact
3	$S_3 = +T$	$+(kD)^2/A$	Hypothèse implicite : les recouvrements d'ordre supérieur à deux sont égaux à $(-1)^{r+1}(kD)^r/A^{r-1}$
⋮	⋮	⋮	
r	$(-1)^{r+1}S_r$	$(-1)^{r+1}(kD)^r/A^{r-1}$	

Il convient aussi de noter que la formule de définition de la Couverture nette comporte exactement m termes, alors que la formule d'Agostini en comporte une infinité. Toutefois, si k est inférieur à A/D, la série géométrique converge et l'erreur de troncation sera inférieure à $\pm (kD)^{m-1}/A^{m-2}$

Nous avons terminé notre analyse de la formule d'Agostini : les divers résultats présentés montrent qu'il s'agit d'un modèle non stochastique de type agrégé. Les supports ne sont pas traités séparément puisque la couverture nette ne dépend que de la somme des audiences, de la somme des duplications et d'une constante arbitrairement définie. L'hétérogénéité des individus composant la population n'est pas envisagée.

Sur le plan empirique, les performances du modèle dépendent entièrement de l'écart existant entre la valeur réelle de k pour les données utilisées et la valeur k = 1,125 de la formule d'Agostini. Les résultats détaillés du test du modèle contre nos deux banques de données sont présentées page 397 . Ils font apparaître un pourcentage moyen d'erreurs important, variant entre 14% et 25%, pour les données de Marc dont la valeur k, calculée selon la formule (11) est de 1,607. Par contre, pour les données de Politz, les résultats sont plus satisfaisants puisque le pourcentage moyen d'erreurs varie de 4 à 5 %. Dans tous les cas, la formule d'Agostini conduit à une surestimation systématique de la couverture nette. De plus, cette surestimation s'accentue avec le nombre de supports inclus et avec l'accroissement du rapport D/A.

Sur le plan pratique, la formule d'Agostini présente le grand avantage de ne nécessiter que fort peu de calculs. En effet, pour m supports, elle ne demande que $(m^2 + m)/2$ additions, 1 multiplication et 1 division.

12. La formule d'Hofmans

Nous avons déjà montré dans l'expression (6) que si l'on compare, dans le cas de 2 supports, le modèle d'Agostini avec la définition de la couverture nette, il est possible d'exprimer k en fonction des A_i et des A_{ij}.

$$k_{ij} = \frac{(A_i + A_j)}{(A_i + A_j - A_{ij})} = \frac{\text{Audience Brute de la paire i,j}}{\text{Couverture Nette de la paire i,j}}$$

En remplaçant la constante k d'Agostini, par la variable k_{ij}, on obtient la formule d'Hofmans (11).

$$(17) \qquad R_m = (\Sigma_{i=1}^{m} A_i)^2 / (\Sigma_{i=1}^{m} A_i + \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} K_{ij} A_{ij})$$

Cette transformation revient en fait à calculer une nouvelle constante K pour chaque combinaison de supports. Il est facile de montrer que K n'est autre que la moyenne pondérée des k_{ij} pour chaque paire de supports, les pondérations étant égales aux A_{ij} :

$$K = \frac{\Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m-1} k_{ij} A_{ij}}{\Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}} = \text{moyenne pondérée}$$

D'où l'on tire :

$$KD = K \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_j = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} k_{ij} A_{ij}$$

Ce nouveau modèle supprime la difficulté liée à la constance de k et permet la prise en considération de chacune des duplications entre paires de supports. Néanmoins, la valeur estimée de k ne sera pas en général égale à la valeur exacte donnée par l'équation (11), sauf pour les combinaisons de deux supports. Tout comme le modèle d'Agostini, le modèle d'Hofmans peut s'écrire sous la forme d'une série géométrique infinie alternée, de base A et de raison KD/A.

$$R_m = \frac{A}{1 + KD/A} = A - KD + \frac{(KD)^2}{A} \quad \ldots (-1)^r \frac{(KD)^r}{A^{r-1}}$$

Cette série ne converge que si $KD < A$ ou encore $K < A/D$

Sur le plan théorique le modèle d'Hofmans demeure un modèle non stochastique et agrégé ne permettant pas l'estimation de la distribution des occasions de voir.

Sur le plan empirique les estimations du modèle d'Hofmans sont supérieures à celles du modèle d'Agostini, bien que la tendance à la surestimation de la couverture nette persiste. On observe également un léger accroissement du taux de surestimation lorsque le nombre de supports s'accroît. Le pourcentage moyen d'erreurs varie de 2,1% à 6,0% pour les données de Marc et de 1,5% à 2,5% pour les données de Politz.

Le modèle d'Hofmans est donc relativement satisfaisant sur le plan empirique. Son plus grave défaut est de ne pas permettre l'estimation de la distribution des occasions de voir.

Sur le plan pratique, le volume des calculs s'accroît par rapport au modèle d'Agostini, sans pour autant devenir prohibitif.

13. La formule de Starch

Daniel Starch (15) a développé une formule simple permettant d'estimer la couverture nette d'une combinaison de supports à partir des audiences et des duplications par paires. Cette formule est du type Ad Hoc puisqu'elle ne repose sur aucune hypothèse explicite concernant les comportements d'exposition. Elle ne permet pas non plus d'estimer la distribution des occasions de voir. Plutôt que de tenter d'estimer directement la couverture nette d'une combinaison de m magazines, la formule de Starch estime le gain net d'audience obtenu en ajoutant un nouveau support à ceux déjà inclus dans la combinaison. Si nous notons $\Delta R_{i/n}$ le gain net d'audience obtenu par l'addition d'un ième support aux n supports déjà inclus dans la combinaison, le modèle de Starch est défini par l'équation suivante :

$$\Delta R_{i/n} = \left(\frac{A_i - A_{in}}{A_i}\right) * \left(\frac{A_i - A_{in-1}}{A_i}\right) \ldots \left(\frac{A_i - A_{i1}}{A_i}\right) * A_i$$

Lorsque les m supports considérés ont été introduits dans la combinaison, la couverture nette est alors égale à la somme des accroissements d'audience.

$$R_m = \Sigma_{i=1}^{m} \Delta R_{i/n}$$

Il est facile de vérifier que la formule de Starch est exacte pour le cas de deux supports.

$$R_{2/1} = \left(\frac{A_2 - A_{12}}{A_2}\right) * A_2 = A_2 - A_{12} \text{ et } R_2 = A_1 + A_2 - A_{12}$$

Par contre, elle n'est pas, en général, valide dans le cas de 3 supports, comme nous allons le démontrer. Pour cela il suffit de comparer les estimations de la formule de Starch avec la formule de définition de la couverture nette. Dans le cas de 3 supports la couverture nette est égale à :

$$R = A_1 + A_2 + A_3 - A_{12} - A_{13} - A_{23} + A_{123}$$

Par contre, la formule de Starch nous donne :

$$R = A_1 + A_2 - A_{12} + (A_3 - A_{13})\ (A_3 - A_{23})\ /\ A_3$$

Ce qui peut être réécrit sous la forme suivante :

$$R = A_1 + A_2 + A_3 - A_{12} - A_{13} - A_{23} + A_{13} * A_{23}/A_3$$

En comparant cette équation avec l'équation de définition de la couverture nette, on observe que la formule de Starch estime la triplication A_{123} par le produit des deux duplications contenant le 3ième support divisé par l'audience du 3ième support. L'hypothèse implicite du modèle de Starch est donc la suivante :

$$A_{123} = A_{13} * A_{23}\ /\ A_3$$

Toutefois, si au lieu d'introduire les supports dans l'ordre 1,2,3, nous les introduisons dans l'ordre 2,3,1, nous obtenons :

$$A_{123} = A_{12} * A_{13}\ /\ A_1$$

De même, si nous introduisons les supports dans l'ordre 3,1,2, nous avons :

$$A_{123} = A_{12} * A_{23}\ /\ A_2$$

Il est donc clair que la formule de Starch n'est valide que si et seulement si :

$$A_{123} = A_{13} * A_{23}\ /\ A_3 = A_{12} * A_{13}\ /\ A_1 = A_{12} * A_{23}\ /\ A_2$$

Rien ne garantit qu'il en soit ainsi et les tests effectués sur nos données conduisent à rejeter cette hypothèse implicite du modèle de Starch. De plus, lorsque l'hypothèse précédente n'est pas vérifiée, les estimations de la formule de Starch varient selon l'ordre dans lequel les supports entrent dans la combinaison, ce qui est, bien entendu, inacceptable. L'exemple numérique suivant montre l'importance de ces variations.

Exemple Numérique

Supposons que nous voulions estimer la couverture nette des trois supports A, B, C, à partir des données de Marc (7).

Supports	Audiences	Paires	Audiences Dupliquées
A	225	AB	183
B	285	AC	181
C	300	BC	228

Par application de la formule de Starch, nous obtenons :

$$R_1 = A = 225$$

$$R_{2/1} = (B-AB) = 102$$

$$R_{3/12} = \left[(C-AC)/C\right] * \left[(C-BC)/C\right] * C = 29$$

$$R_3 = R_1 + R_{2/1} + R_{3/12} = 356$$

La couverture nette réelle de la combinaison est en fait de 372. La formule de Starch, dans l'ordre A, B, C conduit à une sous estimation de 16 unités, soit - 4,3%. Par contre, l'estimation dans l'ordre C, A, B, donne une couverture nette estimée de 364, soit une erreur de - 2,15%. Enfin, l'ordre C, B, A donne une couverture nette estimée de 365 soit une erreur de - 1,9%.

Puisque la couverture nette estimée du modèle de Starch dépend de l'ordre d'inclusion des supports, il est possible d'améliorer les estimations en recherchant l'ordre d'inclusion optimal qui minimise l'erreur d'estimation.

Divers tests empiriques semblent indiquer que le support possédant le plus grand rapport - audience exclusive/audience propre - doit être le dernier inclus. L'avant-dernier support sera celui des supports restant qui possède le plus grand rapport - audience exclusive/audience propre - , et ainsi de suite jusqu'à équisement des m magazines.

L'application de cette règle améliore sensiblement la précision des estimations, spécialement pour les combinaisons comportant un grand nombre de supports. Toutefois, la mise en oeuvre de la règle d'ordonnancement des inclusions accroît le volume des calculs au point de les rendre impraticables pour les combinaisons incluant un nombre élevé de supports.

Pour les données de Marc, la formule de Starch conduit à un pourcentage moyen d'erreurs variant entre + 1,8% et - 6,3%. Il semble que le pourcentage d'erreurs s'accroît en fonction du nombre de supports inclus, puisque pour les données de Politz l'erreur s'élève de - 6% dans le cas de 5 supports à - 21% dans le cas de 11 supports. La modification de la formule de Starch que nous venons de proposer apporte une amélioration sensible, spécialement lorsque le nombre de supports n'est pas trop élevé. Le pourcentage moyen d'erreurs ne varie plus que de + 1% à - 2,1% pour les données de Marc et de - 4,4% à - 16,6% pour les données de Politz.

La table suivante permet une comparaison détaillée, combinaison de supports par combinaison de supports, des performances de la formule de Starch avec et sans modification.

Performances de la formule de Starch, avec ou sans ordonnancement

Combinaison de Supports	Couverture Nette Observée	Couverture Nette Estimée			
		Starch	% Erreur	Starch avec ordre	% Erreur
Données de Marc					
3 Supports					
A-B-C	372	656	-4,30	365	+1,88
A-B-D	354	343	-3,11	354	0,00
A-B-E	349	340	-2,58	344	-1,43
A-C-D	365	361	-1,09	367	+0,55
A-C-E	356	351	-1,40	351	-1,40
A-D-E	329	332	-0,91	332	-0,91
B-C-D	371	366	-1,35	372	+0,26
B-C-E	364	361	-0,82	361	-0,82
B-D-E	349	344	-1,43	345	-1,15
C-D-E	345	349	+1,15	349	+1,15
4 Supports					
A-B-C-D	382	360	-5,75	379	-0,78
A-B-C-E	378	357	-5,55	357	-5,55
A-B-D-E	366	351	-4,10	361	-1,37
A-C-D-E	370	365	-1,35	365	-1,35
B-C-D-E	374	368	-1,60	368	-1,60
5 Supports					
A-B-C-D-E	384	360	-6,25	379	-1,30
Données de Politz					
5 Supports					
Adultes 18 et +	74 630	70 148	-6,01	70 645	-5,34
Femmes 18 et +	40 710	37 519	-5,11	39 013	-4,17
Hommes 18 et +	33 920	32 518	-4,13	32 703	-3,59
11 Supports					
Adultes 18 et +	83 180	65 254	-21,55	69 632	-16,29
Femmes 18 et +	47 300	36 748	-22,31	39 883	-15,68
Hommes 18 et +	35 880	31 768	-11,46	31 768	-11,46
Age 18 - 34	30 800	25 427	-17,44	25 427	-17,44
Age 35 - 49	25 610	18 487	-27,81	21 357	-16,61
Age 50 et +	26 770	20 175	-24,63	20 870	-22,04

Sur le plan empirique, le modèle de Starch sous-estime systématiquement la couverture nette et cette sous-estimation s'aggrave lorsque le nombre de supports inclus s'accroît. Dans l'ensemble, les performances du modèle de Starch sont bien inférieures à celles du modèle d'Hofmans.

Enfin, la formule de Starch, tout comme celle d'Agostini et d'Hofmans, ne permet pas l'estimation de la distribution des occasions de voir.

Sur le plan pratique, on note que le volume des calculs s'accroît rapidement avec le nombre de supports inclus.

14. Le Modèle de Young

Il s'agit d'un modèle de type Ad Hoc, développé spécifiquement pour estimer la couverture nette des stations de radio par L.F. Young (16). Young utilise une simple analyse de régression multiple dans laquelle la couverture nette est une fonction de l'audience brute, du nombre de stations utilisées, du nombre de tranches horaires programmées et du nombre de spots publicitaires par station. Il observe que la fonction donnant le meilleur ajustement est une exponentielle modifiée :

$$(19) \qquad R = P * \left[1 - A(\exp(-\Sigma_{i=1}^{n} b_i x_i))\right]$$

avec R = Couverture nette en proportion de la population cible

A = Constante à estimer

X_1 = Audience brute

X_2 = Nombre de stations

X_3 = Nombre de tranches horaires

X_4 = Nombre de spots par station

Les paramètres b_i de la régression sont estimées par une transformation logarithmique des données :

$$\text{Log}\ (1 - R/P) = \text{Log}\ A - b_1 X_1 - b_2 X_2 - b_3 X_3 - b_4 X_4$$

Le coefficient de corrélation multiple obtenu varie entre R^2=0,865 et R^2=0,95 selon la définition de la population cible.

On observe que la couverture nette s'accroît lorsque les 3 premières variables s'accroissent et diminue quand le nombre de spots par station s'accroît, ce qui est conforme aux expectatives. Néanmoins, la validité d'un modèle de régression est toujours sujette à caution dès lors qu'on souhaite l'appliquer à des situations autres que celle qui a servi à en estimer les paramètres. Les paramètres de la régression reflètent nécessairement les caractéristiques des situations régionales et des stations de radio qui ont été incluses dans l'échantillon, aussi bien que la période de l'année pendant laquelle les données ont été collectées. Il n'est donc pas possible de généraliser les résultats dans le temps ou dans l'espace, et a fortiori pour un autre média. Comme tous les modèles de type Ad Hoc, le modèle de Young ne permet pas d'estimer la distribution des occasions de voir.

Nous allons passer maintenant à l'étude d'un modèle de duplication proposé par Little et Lodish (17) et qui sert de base au modèle de média planning MEDIAC.

15. Le Modèle de Little et Lodish

Le modèle proposé par les auteurs de Mediac est en réalité un modèle mixte combinant une approche stochastique avec l'intervention d'une constante de duplication déterminée empiriquement.

Soit P_i la probabilité d'exposition au support i et P_{ij} la probabilité d'exposition simultanée aux supports i et j. Les valeurs de P_i et P_{ij} peuvent être estimées par $P_i = A_i/POP$ et $P_{ij} = A_{ij}/POP$.

Dans l'hypothèse d'indépendance des audiences, la probabilité d'exposition simultanée aux supports i et j est donnée par :

$$P'_{ij} = P_i * P_j$$

Little et Lodish assument l'existence d'une relation fonctionnelle entre la probabilité empiriquement observée P_{ij} et la probabilité théorique P'_{ij} de la forme : $P_{ij} = (P_i * P_j)^b$.

Utilisant les données de Simmons pour l'année 1967, ils trouvent que la valeur b = 0,65 permet le meilleur ajustement.

L'équation (19) est similaire à la formule d'Agostini en ce sens que la constante b=0,65 s'applique de façon indiscriminée à toute combinaison de supports, comme c'était le cas pour la constante k d'Agostini.

Cette constante b n'est rien d'autre qu'une estimation du degré moyen de corrélation entre les audiences des supports inclus dans la banque de donnée utilisée par Little et Lodish. Nous pouvons donc prévoir a priori que les estimations produites par la formule de Litte et Lodish seront inférieures à celles de la formule d'Hofmans qui utilise les duplications observées entre chaque paire de supports, plutôt qu'une moyenne estimée une fois pour toutes.

Les auteurs créditent leur formule d'une erreur absolue moyenne de 10,7% et d'une erreur moyenne de 1,7%. Ces résultats sont comparables à ceux de la formule d'Agostini, mais ils ne sont pas meilleurs. Little et Lodish défendent leur formule en faisant valoir qu'elle supprime la nécessité de mettre en mémoire dans l'ordinateur toutes les duplications par paires qui sont utilisées par la formule d'Hofmans.

Quoiqu'il en soit, la formule de Little et Lodish ne peut être utilisée que pour estimer la couverture nette de deux supports :

$$R_2 = POP * \left[p_i + p_j - (p_i * p_j)^{0,65}\right] \qquad (20)$$

Si l'on veut estimer la couverture nette de 3 supports ou plus, il faut estimer, d'une manière ou d'une autre, les triplications et les recouvrements d'ordre supérieur à deux, à partir des audiences et des duplications qui sont les seules données disponibles. Little et Lodish tentent de résoudre ce problème en s'efforçant d'établir une relation entre les moments d'ordre supérieur à deux et les deux premiers moments de la distribution des occasions de voir, qui, comme nous l'avons vu, peuvent être estimés à partir de la somme des audiences et de la somme des duplications.

Toutefois, Little et Lodish reconnaissent n'être point parvenus à découvrir une loi de probabilité qui reproduise fidèlement les données tout en restant suffisamment simple (19). Ils se rabattent alors sur une approximation linéaire de la relation entre les moments d'ordre supérieur et les 2 premiers moments. Cependant, comme Little et Lodish ne divulguent pas les équations de ces approximations linéaires il est impossible d'en tester la validité. Nous

devons donc, faute de mieux, nous contenter de tester la formule d'approximation des P_{ij}.

Nous utiliserons pour cela les données de Politz, mais auparavant nous allons donner un exemple numérique d'utilisation de la formule de Little et Lodish.

Exemple numérique

Supposons que nous cherchions à estimer la couverture nette de Better Home & Gardens et de Family Circle, parmi la population adulte. Les données d'audience, en millions d'individus, fournies par Politz sont les suivantes :

B.H.G = 18 770 ; Family Circle = 14 270 ; Population = 120 360

Les probabilités d'exposition aux deux supports sont donc respectivement égales à :

$$p_1 = 0{,}15595 \quad ; \quad p_2 = 0{,}11856$$

En utilisant la formule (20), la couverture nette estimée est égale à :

$$R_2 = 120\ 360 \left[0{,}15595+0{,}11856-(0{,}15595*0{,}11856)^{0{,}65}\right]$$

$$R_2 = 24\ 145$$

La couverture nette réelle mesurée par Politz étant de 28 040, la formule de Little et Lodish conduit à une sous-estimation de 3 895 milliers d'individus, soit 13,9%. Ce pourcentage d'erreurs extrêment élevé ne constitue pas un cas particulier. Les 3 tables suivantes, qui comparent les estimations du modèle de Little et Lodish aux couvertures nettes réelles, pour 55 paires de supports et pour 3 segments de population, montrent que les pourcentages d'erreurs sont clairement inacceptables.

TEST DU MODÈLE DE LITTLE ET LODISH SUR LA BASE DE LA COUVERTURE NETTE DE 55 PAIRES DE MAGAZINES*
(pour chaque paire de magazines, l'erreur absolue de prédiction est donnée par la valeur au-dessus de la diagonale, tandis que l'erreur relative en pourcentage est donnée par la valeur au-dessous de la diagonale)

	B.H.G.	Fam. Cir.	G.HSKP.	L.H.J.	Life	Look	Mc. Call	R. Digest	Redbook	S. Post	Wom. Day
A. INDIVIDUS AGÉS DE 18 ANS OU PLUS											
B.H.G.		- 1 994,82	- 3 330,14	- 3 686,74	- 8 796,76	- 4 214,86	- 4 485,79	- 8 388,28	- 3 686,24	- 6 085,13	- 3 971,16
Fam. Cir.	- 79,89		- 2 380,14	- 2 647,71	- 7 831,78	- 6 689,81	- 3 303,82	- 8 794,20	- 3 895,74	- 6 020,04	- 1 642,77
G.HSKP.	- 56,83	- 44,01		- 1 468,91	- 8 874,84	- 8 927,83	- 2 438,93	- 8 724,48	- 2 711,17	- 5 472,41	- 2 601,85
L.H.J.	- 64,31	- 52,02	- 33,10		- 7 117,45	- 6 692,52	- 2 488,83	- 6 787,42	- 2 787,40	- 4 416,79	- 3 045,03
Life	- 148,09	- 141,80	- 120,28	- 128,01		- 6 687,47	- 8 582,56	- 9 493,15	- 6 516,65	- 7 470,87	- 6 963,94
Look	- 137,60	- 119,81	- 98,82	- 117,81	- 52,91		- 7 510,82	10 445,27	- 6 702,76	- 7 642,98	- 6 912,72
MC. Call	- 62,04	- 50,83	- 41,44	- 32,44	- 114,17	- 44,68		- 7 983,45	- 2 929,10	- 7 044,27	- 1 708,44
R. Digest	- 193,30	- 96,78	- 40,49	- 91,39	- 76,85	- 92,11	- 79,75		- 6 482,24	- 8 165,12	- 6 675,59
Redbook	- 114,65	- 133,43	- 85,28	- 89,81	- 228,41	- 168,22	- 63,62	- 165,16		- 5 038,63	- 2 947,16
S. Post	- 99,11	- 142,99	- 109,84	- 47,80	- 86,42	- 91,18	- 112,08	- 78,96	- 174,88		- 5 547,01
Wom.Day	- 93,88	- 29,00	- 59,00	- 96,44	- 162,10	- 145,41	- 24,15	- 112,76	- 127,58	- 146,70	
B. POPULATION FÉMININE DE 18 ANS OU PLUS											
B.H.G.		- 3 599,17	- 3 099,58	- 3 033,46	- 4 885,10	- 5 014,87	- 3 879,87	- 5 512,87	- 3 393,08	- 3 769,14	- 3 678,17
Fam. Cir.	- 63,10	-	- 2 457,18	- 2 644,90	- 3 978,09	- 3 991,80	- 3 167,85	- 4 543,56	- 3 102,00	- 3 606,45	- 1 886,18
G.HSKP.	- 63,81	- 50,87		- 2 001,20	- 3 724,31	- 3 542,83	- 2 528,23	- 3 865,10	- 2 812,74	- 3 245,25	- 2 583,72
L.H.J.	- 63,86	- 58,91	- 37,85		- 3 887,64	- 3 805,77	- 2 508,88	- 4 209,13	- 2 565,03	- 2 632,39	- 3 124,87
Life	- 118,00	- 92,40	- 78,74	- 86,43		- 3 379,19	- 4 432,98	- 4 763,36	- 4 124,81	- 3 574,95	- 4 185,03
Look	- 124,44	- 92,82	- 71,86	- 84,76	- 84,15		- 4 026,88	- 5 989,06	- 3 668,81	- 3 723,14	- 3 892,15
MC. Call	- 63,60	- 52,97	- 37,07	- 37,64	- 71,73	- 60,91	-	5 154,00	- 2 810,97	- 4 313,08	- 3 921,38
R. Digest	- 99,87	- 81,57	- 62,25	- 71,22	- 68,10	- 103,80	- 65,91		- 4 047,43	- 4 642,52	- 4 593,61
Redbook	- 129,51	- 149,41	- 107,09	- 86,93	- 181,71	- 133,90	- 66,77	- 107,36		- 3 313,23	- 4 593,61
S. Post	- 93,30	- 101,65	- 79,74	- 88,11	- 75,58	- 80,94	- 88,36	- 84,26	- 149,24		- 3 552,11
Wom.Day	- 100,23	- 38,89	- 60,09	- 86,70	- 113,53	- 98,79	- 83,08	- 92,61	- 132,45	- 110,66	
C. POPULATION MASCULINE DE 18 ANS OU PLUS											
B.H.G.		- 757,86	- 600,37	- 1 016,68	- 3 441,68	- 2 904,27	- 1 069,68	- 2 739,84	- 565,38	- 2 067,13	- 614,93
Fam. Cir.	- 92,36		- 472,81	- 511,15	- 2 398,98	- 1 856,44	- 822,46	- 1 808,40	- 445,17	- 1 883,19	- 247,05
G.HSKP.	- 69,44	- 92,71		- 484,42	- 2 275,95	- 1 848,34	- 1 110,56	- 2 306,61	- 259,89	- 1 686,43	- 622,66
L.H.J.	- 124,78	- 88,19	- 74,53		- 2 572,46	- 2 292,52	- 689,55	- 2 251,05	- 633,90	- 1 865,51	- 427,67
Life	- 191,20	- 299,87	- 229,89	- 229,68	-	- 3 304,52	- 3 141,10	- 4 667,94	- 1 683,19	- 3 891,23	- 1 752,83
Look	- 149,70	- 168,77	- 157,98	- 204,69	- 60,68		- 2 782,87	- 4 437,00	- 1 515,54	- 3 970,49	- 1 732,17
MC. Call	- 94,66	- 158,16	- 427,14	- 61,41	- 237,96	- 207,68		- 2 353,95	- 633,32	- 2 136,57	- 519,96
R. Digest	- 105,35	- 124,72	- 226,14	- 149,08	- 75,78	- 79,66	- 107,49		- 1 976,71	- 3 508,20	- 1 487,09
Redbook	- 95,83	- 171,20	- 56,50	- 352,17	- 255,03	- 233,16	- 180,95	- 482,13		- 1 282,84	- 455,23
S. Post	- 98,43	- 285,33	- 185,32	- 174,35	- 89,33	- 103,90	- 151,53	- 68,79	- 221,18		- 1 324,35
Wom.Day	- 109,81	- 52,56	- 563,51	- 106,92	- 278,23	- 368,55	- 108,52	- 152,82	- 650,33	- 232,34	

* Données : POLITZ MAGAZINE STUDY, 1964.

Un tel résultat était d'ailleurs prévisible, eu égard aux hypothèses du modèle. En effet, sur le plan théorique, nos auteurs assument à la fois l'homogénéité des individus, puisque p_i est identique pour tous les individus, et l'homogénéité des duplications puisque le coefficient d'ajustement b est invariant quelle que soit la paire de supports considérés.

Avant de passer en revue les modèles stochastiques, on peut dresser le bilan suivant des modèles empiriques.

1 - Ils n'expliquent pas le phénomène de duplication des audiences, mais se contentent de l'approximiser.

2 - Ils ne permettent pas d'estimer la distribution des occasions de voir.

3 - Ils ne procurent une approximation satisfaisante de la couverture nette que lorsqu'ils prennent en compte l'ensemble des duplications entre paires de supports et non leur moyenne.

2. Les Modèles Stochastiques

Tous ces modèles définissent l'exposition aux supports comme un phénomène aléatoire régit par une certaine loi de probabilité. Par contre, ils diffèrent quant aux hypothèses faites concernant la présence ou l'absence d'homogénéité entre les individus, la présence ou l'absence d'homogénéité entre les supports et enfin la présence ou l'absence d'indépendance entre les audiences des divers supports.

Il est donc possible de les classer, du plus simple au plus complexe, selon la nature des hypothèses faites; une telle classification est présentée dans le tableau suivant :

Classification des Modèles Stochastiques de Duplication d'Audience

Hypothèses			Modèles
Les individus sont :	Les supports sont :	Les audiences sont :	
1 - Homogènes	Homogènes	Indépendantes	Binomial
2 - Homogènes	Homogènes	Interdépendantes	Kwerel
3 - Homogènes	Hétérogènes	Indépendantes	Sainsbury
4 - Homogènes	Hétérogènes	Interdépendantes	Sainsbury Modifié
			Agrégation Séquentielle
			Indépendance Conditionnelle
			Hofmans Etendu
5 - Hétérogènes	Hétérogènes	Interdépendantes	Metheringham
6 - Hétérogènes	Hétérogènes	Interdépendantes	Benson

Les trois hypothèses fondamentales concernant les individus, les supports et les audiences peuvent être définies de la façon suivante :

- L'hypothèse d'homogénéité des individus implique que tous les membres de la population cible possèdent la même probabilité d'exposition vis-à-vis d'un support donné :

$$P_{a,i} = P_{b,i} = \ldots = P_i \ \forall\ a,b \in POP, \forall\ i \in M$$

En d'autres termes, cette hypothèse réduit la distribution des probabilités individuelles d'exposition à un point unique P_i de masse I et autorise l'agrégation des individus.

- L'hypothèse d'hétérogénéité des individus signifie que l'on postule l'existence d'une certaine distribution des probabilités individuelles d'exposition à un support, telle la loi Bêta par exemple.

- L'hypothèse d'homogénéité des supports signifie que la probabilité d'exposition d'un individu donné est invariante quel que soit le support considéré :

$$P_{a,i} = P_{a,j} = \dots = P_a \quad \forall\, i, j \in M, \forall\, a \in POP$$

et que, par conséquent, on s'autorise à agréger les supports.

- L'hypothèse d'hétérogénéité des supports implique que :

$$P_{a,i} \neq P_{a,j} \;\forall\, i,j \in M,\; i \neq j,\; \forall\, a \in POP$$

et qu'il convient d'estimer séparément chacune de ces probabilités.

- L'hypothèse d'indépendance des audiences signifie que l'exposition aux supports constitue un ensemble d'évènements mutuellement indépendants et tels que l'exposition à un support ne modifie en rien la probabilité d'exposition à un autre support :

$$P(A_i/A_j) = P(A_i) \quad \forall\, i,j \in M,\; i \neq j$$

$P\,(A_{ij}) = P(A_i) * P(A_j)$ Il y a $\binom{m}{2}$ égalités de ce type.

$P\,(A_{ijk}) = P(A_i) * P(A_j) * P(A_k)$ Il y a $\binom{m}{3}$ égalités de ce type.

$\vdots$ $\qquad\qquad\qquad\qquad$ $\vdots$

$P(A_{12\dots m}) = P(A_1) * P(A_2) * \dots * P\,(A_m)$ Une égalité seulement.

- L'hypothèse d'interpédendance des audiences signifie soit qu'aucune des égalités précédentes n'est vérifiée, il s'agit alors d'interdépendance globale, soit que seulement certaines égalités précédentes ne sont pas vérifiées, et il s'agit alors d'une interdépendance conditionnelle.

Il convient de remarquer que nous n'obtenons que 6 familles de modèles et non pas les 8 attendues résultant de la classification croisée de 3 critères dichotomiques, car l'hypothèse d'interdépendance des audiences devient absurde quand les individus sont considérés comme hétérogènes.

21. Le Modèle Binomial

Hypothèses : 1 - Les individus sont homogènes
2 - Les supports sont homogènes
3 - Les audiences dont indépendantes

Par la combinaison des 2 premières hypothèses, la probabilité d'exposition devient identique pour tous les individus et tous les supports. Cette probabilité unique sera estimée en divisant l'audience moyenne des m supports considérés par la population :

$$p = \overline{A}/POP \text{ avec } \overline{A} = \frac{1}{m} \Sigma_{i=1}^{m} A_i$$

Chaque support représentant un tirage indépendant d'égale probabilité, la probabilité qu'un individu soit exposé à exactement k supports parmi les m est de :

$$(21) \qquad P\ (k/m,p) = \binom{m}{k} p^k (1 - p)^{m-k}$$

La probabilité qu'un individu soit exposé à au moins 1 support parmi les m est de :

$$P\ (k \geqslant 1/m,p) = 1 - P\ (k = o/m,p) = 1 - (1 - p)^m$$

Comme tous les individus sont homogènes, le nombre d'individus exposés à au moins un support parmi les m sera de :

$$(22) \qquad R_m = POP * \left[1 - (1 - p)^m \right] = \text{Couverture nette.}$$

Le modèle binomial permet aussi d'estimer la distribution des occasions de voir en multipliant la probabilité de chaque classe de fréquence par la population :

$$(23) \qquad E_k^m = POP * P\ (k/m,p)$$

L'exemple numérique suivant décrit le fonctionnement du modèle binomial de duplication d'audience.

Exemple numérique

Supposons que nous cherchions à estimer la couverture nette de la combinaison des cinq magazines suivants : Life, Look, Mc Call's, Reader's Digest et Saturday Evening Post. Les seules données requises pour l'application du modèle binomial sont les audiences de chacun des magazines et le nombre d'individus constituant la population cible.

Magazines	Audiences	(Politz-1946- Milliers d'Adultes)	
Life	30 500	Population	= 120 360
Look	28 760	Audience brute	= 139 890
Mc Call's	21 430	Audience moyenne	= 27 978
R. Digest	36 320	Probabilité d'Exposition	= 0,23245

Une fois calculée la probabilité moyenne d'exposition, nous obtenons la couverture nette estimée en utilisant l'équation (22) :

$$R_5 = 120\ 360 * \left[1 - (1 - 0{,}23245)^5 \right] = 88\ 297$$

La couverture nette réelle, mesurée par Politz, n'est que de 74 630. Le modèle binomial conduit donc à une surestimation de 13 667 milliers d'individus, soit 18,3% ! Il ne s'agit pas là d'un résultat isolé mais bien d'une très forte surestimation systématique, comme le montre la table suivante qui compare la couverture nette estimée avec la couverture nette mesurée par Politz pour 9 segments de population.

Test du Modèle Binomial pour une Combinaison de 5 Magazines et pour 9 Segments de Population

Segment de Population	Milliers d'individus	Couverture Nette observée	Couverture Nette estimée	Erreur Absolue	Pourcentage d'erreurs
Age 10 ans ou +					
Ensemble	149 600	93 530	109 784	16 254	17,38
Femmes	77 790	50 440	59 280	8 840	17,52
Hommes	71 180	43 090	50 291	7 201	16,71
Age 15 ans ou +					
Ensemble	130 140	81 400	96 141	14 741	18,11
Femmes	68 390	44 360	52 588	8 228	18,55
Hommes	61 750	37 040	43 331	6 291	16,98
Age 18 ans ou +					
Ensemble	120 360	74 630	88 297	13 667	18,31
Femmes	63 240	40 710	48 246	7 536	18,51
Hommes	57 120	33 920	39 863	5 943	17,21

Cette table montre que le modèle binomial surestime la couverture nette de 17,7% en moyenne pour cette combinaison de 5 magazines, et que cette surestimation est à peu près identique pour chaque segment de population. Que se passe-t-il quand le nombre de supports inclus dans la combinaison s'accroît? La table suivante répond à cette question dans le cas de 11 magazines.

Test du Modèle Binomial pour une Combinaison de 11 Magazines et pour 9 segments de Population

Segment de Population	Milliers d'individus	Couverture Nette observée	Couverture Nette estimée	Erreur Absolue	Pourcentage d'erreurs
Age 10 ans ou +					
Ensemble	149 600	103 610	129 576	25 966	25,06
Femmes	77 790	58 090	71 756	13 666	25,53
Hommes	71 180	45 520	54 949	10 047	22,07
Age 15 ans ou +					
Ensemble	130 140	90 460	113 574	23 114	25,55
Femmes	68 390	51 380	63 488	12 108	23,57
Hommes	61 750	39 080	47 262	8 182	20,94
Age 18 ans ou +					
Ensemble	120 360	83 180	104 758	21 578	25,94
Femmes	63 240	47 300	58 069	11 309	23,91
Hommes	57 120	35 880	43 636	7 756	21,16

On observe que l'accroissement du nombre de supports conduit à un accroissement du degré de surestimation de la couverture nette. Pour 11 magazines, le pourcentage moyen de surestimation est supérieur à 23%. De toute évidence, la modèle binomial est donc clairement inacceptable sur le plan des performances empiriques. Il surestime très considérablement la couverture nette car il sous-estime très largement les duplications. En conséquence, il est aussi incapable de restituer correctement la distribution des occasions de voir.

Sur le plan théorique, le modèle binomial est extrêmement insatisfaisant puisqu'il est basé sur des hypothèses qui sont toutes trois contredites

par les statistiques d'audience. Son intérêt est surtout de constituer un point de référence contre lequel juger les performances des autres modèles.

Sur le plan pratique, malgré la très mauvaise qualité des estimations, l'extrême simplicité du modèle binomial a poussé certains auteurs à en recommander l'utilisation (20). Par ailleurs, si l'on ne dispose que des audiences de chaque support et que les duplications sont inconnues, c'est le seul modèle applicable puisque tous les autres modèles, à l'exception du modèle de Sainsbury, nécessitent la connaissance des duplications de chaque paire de supports.

22. Le Modèle de Sainsbury (6)

Hypothèses : 1 - Les individus sont homogènes
2 - Les supports sont hétérogènes
3 - Les audiences sont indépendantes

Ce modèle s'obtient par la relaxation de l'hypothèse d'homogénéité des supports du modèle binomial. Chaque support est désormais caractérisé par une probabilité d'exposition propre : $P_i = A_i/POP$. Dans ce cas, la probabilité qu'un individu soit exposé à exactement k supports parmi les m est donnée par l'application du théorème de Waring (14).

$$P(k/m, P_1, P_2, \ldots, P_m) = S_k - \binom{k+1}{k} S_{k+1} + \binom{k+2}{k} S_{k+2} \ldots (-1)^{m-k} S_m$$

(24) $$P(k/m, \ldots P_i \ldots) = \Sigma_{i=k}^{m} (-1)^{i-k} \binom{i}{k} S_i; \quad \leqslant k \leqslant m; \; S_o = 1$$

La probabilité qu'un individu soit exposé à au moins 1 support est égale à :

$$P(k \geqslant 1/m, \ldots P_i \ldots) = 1 - P(k = o/m, \ldots P_i \ldots)$$
$$= S_1 - S_2 + S_3 \ldots \pm S_m$$

Et, comme les individus sont homogènes, la couverture nette est égale à :

(25) $$R_m = POP * \left[S_1 - S_2 + S_3 - \ldots \pm S_m \right]$$

En vertu de l'hypothèse 3, la probabilité d'un recouvrement d'ordre k est égale au produit des probabilités d'exposition à chacun des k supports inclus dans le recouvrement considéré :

$$P(A_{ij \ldots k}) = P_i * P_j * \ldots * P_k$$

Il en résulte que la somme S_k des recouvrements d'ordre k, est donné par :

$$S_k = \sum_{i \neq j} \sum \ldots \sum_{\neq k} P(A_{ij \ldots k})$$

Comme pour le modèle binomial, la distribution des occasions de voir s'obtient aisément à partir des probabilités de chaque classe de fréquence par :

$$(25) \qquad E_k^m = POP * [\, P\,(k/m, \ldots P_i \ldots)\,]$$

Sur le plan de la conduite des calculs, lorsque l'on désire estimer uniquement la couverture nette, il est préférable de remplacer l'équation (25) par la formule suivante qui nécessite nettement moins de calculs.

$$(26) \qquad R_m = POP * [\, 1 - (1-P_1)\,(1-P_2) \ldots (1-P_m)\,]$$

Toutefois, la formule (26) ne fait pas apparaître le lien existant entre l'estimation de la couverture nette et l'estimation de la distribution des occasions de voir.

L'exemple numérique suivant décrit la conduite des calculs pour une combinaison de 3 magazines.

Exemple numérique

Supposons que nous désirions estimer la couverture nette des 3 magazines suivants, d'après les données de l'enquête Politz de 1964.

Magazines	Audiences	Population
Better Homes & Gardens	18 770	Age 18 Ans et + = S_0 = 120 360
Family Circle	14 270	Somme des audiences = S_1 = 47 790
Good Housekeeping	14 750	

Nous calculons tout d'abord les duplications et la triplication par application de l'hypothèse d'indépendance des audiences.

Paires de supports	Duplication	Somme des Duplications
B.H.G - F.C.	18 770 * 14 270/120 360 = 2 235	S_2 = 6 274
B.H.G.- G.H.	18 770 * 14 750/120 360 = 2 300	
F.C. - G.H.	14 270 * 14 750/120 360 = 1 749	

Triplication

B.H.G. - F.C. - G.H. = 18 770 * 14 270 * 14 750/$(120\ 360)^2$ = 273 = S_3

Nous pouvons maintenant estimer la couverture nette :

$$R_3 = S_1 - S_2 + S_3 = 41\ 789$$

La distribution des occasions de voir s'obtient par application du théorème de Waring.

Nombre d'expositions			Théorème de Waring		Nombre d'individus
0 sur 3	E_0^3	=	$S_0 - S_1 + S_2 - S_3$	=	78 571
1 sur 3	E_1^3	=	$S_1 - 2S_2 + 3S_3$	=	36 061
2 sur 3	E_2^3	=	$S_2 - 3S_3$	=	5 455
3 sur 3	E_3^3	=	S_3	=	273

Les deux tables suivantes comparent les estimations du modèle de Sainsbury avec la couverture nette réelle observée pour une combinaison de 5 magazines et une combinaison de 11 magazines, sur la base des données Politz de 1964.

Modèle de Sainsbury

Estimation de la couverture nette

Combinaison de 5 Magazines : Life, Look, Mc Call's, Digest, Post					
Segment de Population	Nombre d'individus	Couverture Nette Observée	Couverture Nette Estimée	Erreur Absolue	Pourcentage d'Erreurs
Age 10 ans et +					
Ensemble	149 600	93 530	110 109	16 579	17,73
Femmes	77 890	50 440	59 416	8 976	17,80
Hommes	71 810	43 090	50 573	7 663	17,80
Age 15 ans et +					
Ensemble	130 140	81 400	96 441	15 041	18,48
Femmes	68 390	44 360	52 717	8 357	18,84
Hommes	61 750	37 040	43 754	6 714	18,13
Age 18 ans et +					
Ensemble	120 360	74 630	88 572	13 942	18,68
Femmes	63 240	40 710	48 376	7 666	18,83
Hommes	57 120	33 920	40 250	6 330	18,86

Combinaison de 11 magazines : Life, Look, Mc Call's, Digest, Post Better Homes & Gardens, Family Circle, Good House Keeping, Ladies Home Journal, Woman's Day, Redbook					
Segment de Population	Nombre d'individus	Couverture Nette Observée	Couverture Nette Estimée	Erreur Absolue	Pourcentage d'erreurs
Age 10 ans et +					
Ensemble	149 600	103 610	130 322	26 712	25,78
Femmes	77 890	58 090	71 796	13 706	23,59
Hommes	71 810	45 520	56 141	10 621	23,33
Age 15 ans et +					
Ensemble	130 140	90 460	114 193	23 733	26,24
Femmes	68 390	51 380	63 495	12 115	23,58
Hommes	61 750	39 080	48 427	9 347	23,92
Age 18 ans et +					
Ensemble	120 360	83 180	105 316	22 136	26,61
Femmes	63 240	47 300	58 619	11 319	23,93
Hommes	57 120	35 880	44 592	8 712	24.28

Commentaires

Les résultats du modèle de Sainsbury ne sont pas meilleurs que ceux du modèle binomial. On observe une surestimation moyenne de 18,3% pour la combinaison de 5 magazines et de 24,6% pour la combinaison de 11 magazines. En ce qui concerne les données de Marc, les résultats sont encore plus désastreux puisque l'on enregistre une surestimation moyenne de 52,9% pour dix combinaisons de trois supports, qui s'élève ensuite à 73,7% pour 5 combinaisons de 4 supports, pour culminer enfin à 90,6% pour la combinaison de 5 supports.

Ces résultats font apparaître que l'abandon de l'hypothèse d'homogénéité des supports n'apporte aucune amélioration. Comme nous allons le voir bientôt, c'est l'hypothèse de duplication aléatoire qui est responsable de la mauvaise qualité des estimations du modèle binomial et du modèle de Sainsbury. Tous les modèles que nous allons étudier maintenant rejettent l'hypothèse de duplication aléatoire.

23. Le Modèle de Kwerel

Ce modèle, comme tous ceux qui vont suivre, diffère du modèle binomial par l'abandon de l'hypothèse d'indépendance des audiences. Les corrélations entre audiences, révélées par toutes les études d'audience quels que soient les pays ou les méthodes de mesure utilisées, peuvent s'expliquer sur la base des deux phénomènes suivants :

1 - Les personnes qui partagent une sphère d'intérêt commune, telle que le sport ou la mode, ont tendance à s'exposer davantage aux supports qui se spécialisent dans ce domaine qu'aux autres supports. Par exemple, si une personne appartient à l'audience d'un magazine consacré à la mode, sa probabilité d'exposition à un autre magazine de mode sera plus élevée que sa probabilité d'exposition à un magazine sportif. En conséquence, il existe une corrélation positive entre les audiences des supports orientés vers une même sphère d'intérêt et une corrélation négative entre les audiences des supports dirigés vers des sphères d'intérêt différentes.

2 - Les personnes dont le taux d'exposition est supérieur à la moyenne ont tendance à être éclectiques dans le choix des supports, tandis que les personnes dont le taux d'exposition est inférieur à la moyenne tendent à se cantonner dans un type de support déterminé.

Tous les modèles qui vont suivre s'efforcent d'extrapoler l'interdépendance globale des audiences à partir de la connaissance des duplications par paires qui n'étaient pas prises en compte par les deux modèles précédents.

Les hypothèses du modèle de Kwerel (21) sont les suivantes :

Hupothèses : 1 - Les individus sont homogènes
2 - Les supports sont homogènes
3 - Les audiences sont interdépendantes

Le modèle de Kwerel (22), basé sur la théorie du maximum d'entropie (23), montre que la couverture nette de m supports est nécessairement comprise entre les 2 limites suivantes :

(27) Limite Maximum $= \Sigma_{i=1}^{m} A_i - \frac{2}{m} \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}$

(28) Limite Minimum $= (h+1)\ \bar{A} - \frac{1}{2}\ h\ (h+1)\ \bar{D}$

avec h = Minimum [(m-1),k] ; k = plus grand entier contenu dans $\bar{A}/\bar{D}$

et $\bar{A} = \frac{1}{m} \Sigma_{i=1}^{m} A_i$; $\bar{D} = 2\ \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}/m\ (m-1)$

Kwerel suggère ensuite d'estimer la couverture nette par :

(29) $R_m = 2\ Lmax * Lmin/(Lmax+Lmin)$

L'exemple numérique suivant illustre l'application de la méthode de Kwerel à l'estimation de la couverture d'une combinaison de 5 magazines sur la base des données de Politz de 1964

Exemple numérique

Magazines	Audiences	Paires de Supports	Duplications	Paires de Supports	Duplications
Life	30 500	Life-Look	12 760	Look-Digest	11 340
Look	28 760	Life-Mc Call's	7 500	Look-Post	8 440
Mc Call's	21 340	Life-Digest	13 140	McCall's-Digest	10 010
Digest	36 320	Life-Post	9 290	McCall's-Post	6 280
Post	22 880	Look-McCall's	79 950	Digest-Post	10 610

A l'aide des données nous calculons les quatre quantités suivantes :

A = Audience brute = 139 890

$\bar{A}$ = Audience Moyenne = 27 978

D = Somme des Duplications = 97 320

$\bar{D}$ = Duplication Moyenne = 9 732

Nous pouvons maintenant calculer la limite maximum et la limite minimum de la couverture nette à l'aide des équations (27) et (28).

Limite maximum = 139 890 - 2 * 97 320/5 = 100 962

Le rapport de l'audience moyenne à la duplication moyenne est de 2,87. Si nous comparons le plus grand entier contenu dans 2,87, c'est-à-dire 2, avec le nombre de supports moins un, c'est-à-dire 4, la plus petite valeur est : h = 2. En conséquence, la limite minimum de la couverture nette est donnée par :

Limite minimum = 3 * 27 978 - 3 * 9 732 = 54 738

On observe que l'écart entre les deux limites est fort grand. Néanmoins, l'application de la formule (29) nous permet d'estimer la couverture nette :

$$R_5 = 2 * 100\ 962 * 54\ 738/(100\ 962 + 54\ 738) = 70\ 989.$$

La couverture nette réelle mesurée par Politz étant de 74 630, le modèle de Kwerel conduit à une sous-estimation de 3 641 milliers d'individus soit 4,9%. Ce résultat est supérieur à celui obtenu par la formule d'Agostini mais il est inférieur à celui de la formule d'Hofmans.

Kwerel a démontré que l'écart entre la limite maximum et la limite minimum s'accroît parallèlement au nombre de supports inclus et qu'il est d'autant plus grand que le rapport critique, C = D/A, s'approche de la valeur critique (m-1)/4.

Pour notre exemple, C est égal à 0,696°, ce qui est relativement proche de la valeur critique qui est, dans ce cas, égale à l'unité.

La table suivante compare les estimations du modèle de Kwerel à celles du modèle binomial, du modèle d'Agostini et du modèle d'Hofmans, pour une combinaison de 5 magazines et une combinaison de 11 magazines et pour divers segments de population.

Comparaison des Estimations des Modèles de Kwerel, Agostini, Hofmans et Binomial

Segment de Population	Couverture Nette	Pourcentage d'erreur sur la couverture Nette			
		Kwerel	Agostini	Hofmans	Binomial
	Combinaison de 5 Magazines				
Age 18 ans ou +					
Ensemble	74 630	-4,88	5,15	1,58	18,31
Femmes	40 710	-5,11	5,38	0,91	18,51
Hommes	33 920	-2,77	5,23	2,00	17,21
	Combinaison de 11 Magazines				
Age 18 ans ou +					
Ensemble	83 180	-14,26	1,51	2,50	25,94
Femmes	47 300	-11,68	5,99	0,60	23,91
Hommes	35 880	-10,26	2,60	1,33	21,16
Age 18 à 34	30 800	-10,35	5,92	2,78	18,44
Age 34 à 49	25 610	-17,32	-1,29	-5,52	20,13
Age 50 ans ou +	26 770	-16,99	-1,64	-4,97	36,10

Commentaires

Dans l'ensemble, les résultats du modèle de Kwerel ne sont pas satisfaisants, spécialement pour la combinaison de 11 supports. Le modèle sous-estime systématiquement la couverture nette. Le pourcentage moyen d'erreurs est de 4,25% pour 5 supports et de 13,48% pour 11 supports. Les résultats sont nettement moins bons que ceux obtenus par la formule d'Hofmans. Néanmoins, le modèle de Kwerel est nettement préférable au modèle binomial.

Kwerel présente aussi une méthode d'estimation de la distribution des occasions de voir. Notre auteur fait l'hypothèse que la distribution des occasions de voir suit une loi de probabilité géométrique dont le paramètre k est déterminé en fonction de la couverture nette estimée.

$$(30) \qquad E_i^m = A - \bar{D}\left[1 + (1-k) + (1-k)^2 + \dots (1-k)^{m-2}\right]$$

Le paramètre k est obtenu en résolvant l'équation suivante, de degré m-2 :

$$(31) \qquad A - R_m - \bar{D}\left[\Sigma_{i=1}^{m-1} (m-i)(1-k)^{i-1}\right] = 0$$

Quand le nombre de supports est supérieur à 4, il faut recourir à une méthode de résolution itérative, pour trouver la racine positive (1-k), solution de l'équation (31).

L'exemple numérique suivant, utilisant les données de Marc, décrit les calculs conduisant à l'estimation de la distribution des occasions de voir, pour une combinaison de 5 supports.

Exemple numérique

Magazines	Audiences	Paires	Duplications	Paires	Duplications
A	225	AB	183	BD	117
B	285	AC	181	BE	144
C	300	AD	84	CD	115
D	153	AE	120	CE	164
E	183	BC	228	DE	74

Sommes des audiences : A = 1146
Sommes des duplications : D = 1410
Duplication moyenne : $\bar{D}$ = 141

Nous estimons tout d'abord la couverture nette par la formule (29)

$R_5 = 410$

Nous écrivons ensuite l'équation (31), puis nous cherchons sa racine positive (1-k) par itérations successives.

$$1.146 - 410 - 141 \left[4 + 3\,(1-k) + 2(1-k)^2 + (1-k)^3 \right] = 0$$

Racine : $(1-k) = 0{,}325$

Finalement, à l'aide de l'équation (30), nous obtenons la distribution estimée des occasions de voir, que nous pouvons comparer avec la distribution exacte mesurée par Marc. Cette comparaison est présentée dans la table suivante.

Estimation de la Distribution des Occasions de Voir
par le Modèle de Kwerel

Pour une combinaison de 5 supports

	E_1^5	E_2^5	E_3^5	E_4^5	E_5^5	R_5
Observé	50	86	108	100	40	384
Estimé	114	48	101	104	43	410
% d'Erreurs	+ 128,0	- 44,2	- 6,5	+ 4,0	+ 7,5	+ 6,8

L'ajustement n'est guère satisfaisant. Le mode de la distribution estimée se produit au noveau d'une exposition, tandis que le mode observé se produit au niveau de 3 expositions. Le test d'ajustement du Chi Deux s'élève à 100,6, indiquant la mauvaise qualité des estimations du modèle de Kwerel.

Pour tester plus avant le modèle Kwerel, nous avons estimé la couverture nette et la distribution des occasions de voir pour les 15 combinaisons de supports contenues dans les données de Marc. Les résultats détaillés de ces tests apparaissent dans les tables de la page 397 à la page 491.

En ce qui concerne la couverture nette, le pourcentage moyen d'erreurs s'accroît rapidement avec le nombre de supports considérés. Il passe de 1% seulement dans le cas de 3 supports à 3,3% dans le cas de 4 supports et à 6,8% pour la combinaison de 5 supports. Cette surestimation systématique et croissante de la couverture nette est cependant du même ordre de grandeur que celle observée pour le modèle d'Hofmans.

En ce qui concerne la distribution des occasions de voir, le Chi Deux moyen passe de 2,99 dans le cas de 3 supports à 24,47, dans le cas de 4 supports et à 100,6 pour la combinaison de 5 supports. On observe donc que la qualité de l'estimation de la distribution des occasions de voir se détériore rapidement dès que le nombre de supports devient supérieur à 3.

Sur le plan théorique on remarque que ce modèle recourt non seulement à l'agrégation des audiences des supports lors du calcul de A, mais encore à l'agrégation des duplications par paires pour le calcul de D et qu'en conséquence il n'utilise pas toute l'information contenue dans les données d'audience. Par ailleurs ce modèle introduit deux hypothèses supplémentaires, à savoir que le comportement d'exposition aux supports correspond à une situation d'entropie maximum et que la distribution des occasions de voir suit une loi géométrique.

Sur le plan pratique, il convient de noter que l'estimation de la couverture nette ne nécessite pas de lourds calculs. Par contre, le tableau est moins brillant en ce qui concerne l'estimation de la distribution des occasions de voir, car le volume des calculs nécessaires à l'estimation de la racine k s'accroît rapidement avec le nombre de supports utilisés.

24. Le Modèle de Sainsbury modifié

Ce modèle et les 3 modèles suivants partagent les mêmes 3 hypothèses :

1 - Les individus sont homogènes
2 - Les supports sont homogènes
3 - Les audiences sont interdépendantes.

Ils ne diffèrent que dans leur façon de définir l'interdépendance.

Le modèle de Sainsbury modifié (6) utilise le concept d'audience non-couverte qui n'est autre que le complément à la population totale de la couverture nette.

De même, la probabilité de non-couverture est le complément à 1 de la probabilité d'exposition :

$$q_i = 1 - p_i$$

Dans le cas de deux supports nous avons : $q_{ij} = 1 - p_i - P_j + P_{ij}$

Sainsbury définit ensuite la probabilité de non couverture ajustée q'_i comme la probabilité de non couverture qui serait obtenue dans l'hypothèse d'indépendance des audiences.

(32) $$q'_i = q_{ij}/q_j$$

Il convient de noter ici que la valeur de q'_i varie en fonction du support j choisi comme base d'ajustement.

Le modèle calcule ensuite la couverture nette sur la base du théorème de Waring :

(33) $$R_m = POP * [1 - (q'_1 * q'_2 \dots * q'_m)]$$

Et, bien que cela ne soit pas mentionné dans l'article de MM. Caffyn et Sagovsky, le modèle permet aussi l'estimation de la distribution des occasions de voir par l'utilisation du théorème de Waring.

Sur le plan théorique, la faiblesse du modèle est de conduire à des estimations différentes selon le support utilisé comme base d'ajustement. Les auteurs de l'article précité conseillent de choisir le support présentant la plus grande audience.

L'exemple numérique suivant décrit l'application du modèle de Sainsbury modifié à une combinaison de 3 supports.

Exemple numérique

Supposons que nous désirions estimer la couverture nette et la distribution des occasions de voir des 3 magazines suivants, d'après les données de l'enquête Politz de 1964.

Magazines	Audiences	Non-Couverture	Population
Better Homes & Gardens	18 770	101 590	120 360
Family Circle	14 270	106 090	
Good Housekeeping	14 750	105 610	

Paires de Magazines	Audiences Nettes	Non-Couverture	Duplications
B.H.G. - F.C.	28 040	92 320	5 000
B.H.G. - G.H.	27 660	92 700	5 860
F.C. - G.H.	23 680	96 680	5 340

Puisque le magazine Better Homes & Gardens possède la plus forte audience, il sera utilisé pour calculer les probabilités ajustées de non couverture.

$$q'_1 = q_1 = 101\ 590 / 120\ 360 = 0{,}844$$

$$q'_2 = q_{12}/q_1 = 92\ 320 / 101\ 590 = 0{,}909$$

$$q'_3 = q_{13}/q_1 = 92\ 700 / 101\ 590 = 0{,}912$$

La couverture nette peut maintenant être estimée par la formule (33)

$$R_3 = 120\ 360 * [1 - (0{,}844 * 0{,}909 * 0{,}912)] = 36\ 119$$

La distribution des occasions de voir s'obtient par application du théorème de Waring. Nous calculons d'abord la somme des duplications D, puis les triplications, T.

$$D = 5\ 000 + 5\ 860 + 5\ 340 = 16\ 200$$

$$T = R_3 - A + D = 36\ 119 - 47\ 790 + 16\ 200 = 4\ 529$$

Nombre d'Expositions

0 sur 3	E_0^3	$= POP - A + D - T$	$= 84\ 241$
1 sur 3	E_1^3	$= A - 2D + 3T$	$= 28\ 977$
2 sur 3	E_2^3	$= D - 3T$	$= 2\ 613$
3 sur 3	E_3^3	$= T$	$= 4\ 529$

La table suivante compare les estimations du modèle de Sainsbury Modifié à celles du modèle binomial, du modèle d'Agostini, du modèle d'Hofmans et du modèle de Kwerel, pour une combinaison de 5 magazines et une combinaison de 11 magazines, sur la base des données de Politz de 1964.

Comparaison des Estimations des Modèles de Sainsbury Modifié, Kwerel, Hofmans, Agostini et Binomial

Segment de Population	Couverture Nette	Pourcentage d'erreur sur la couverture nette				
		Binomial	Agostini	Hofmans	Kwerel	Sains.Mod.
		Combinaison de 5 magazines				
Age 18 ans ou +						
Ensemble	74 630	18,31	5,15	1,58	-4,88	6,20
Femmes	40 710	18,51	5,38	0,91	-5,11	7,88
Hommes	33 920	17,21	5,23	2,00	-2,77	4,48
		Combinaison de 11 magazines				
Age 18 ans ou +						
Ensemble	83 180	25,94	1,51	2,50	-14,26	13,56
Femmes	47 300	23,91	5,99	0,60	-11,68	14,15
Hommes	35 880	21,16	2,60	1,33	-10,26	9,49
Age 18 à 34	30 800	18,44	5,92	2,78	-10,35	10,36
Age 35 à 49	25 610	20,13	-1,29	-5,52	-17,32	11,35
Age 50 ans ou +	26 770	36,10	-1,64	-4,97	-16,99	17,27

Commentaires

Les résultats du modèle de Sainsbury modifié sont similaires à ceux du modèle de Kwerel, mais de signe inverse. Le modèle de Sainsbury modifié surestime la couverture nette tandis que le modèle de Kwerel la sous-estime. Pour les 2 modèles, le pourcentage d'erreur s'accroît fortement avec le nombre de supports considérés. Bien que les estimations du modèle de Sainsbury modifié soient supérieures à celles du modèle Binomial et du modèle de Sainsbury non modifié, elles sont néanmoins inférieures à celles du modèle d'Agostini et du modèle d'Hofmans.

En ce qui concerne les données de Marc, les résultats détaillés du modèle de Sainsbury sont présentés dans les tables de la page 397 à la page 401. On observe que, pour ces données, le modèle de Sainsbury modifié est inférieur au modèle de Kwerel, aussi bien en termes de couverture nette qu'en termes de distribution des occasions de voir. Le pourcentage moyen d'erreur sur la couverture nette passe de 3,73% en moyenne pour les combinaisons de 3 magazines à 8,75% pour les combinaisons de 4 magazines et il atteint 14,58% pour la combinaison de 5 magazines. Les estimations de la distribution des occasions de voir sont très mauvaises. Le Chi Deux moyen passe de 34,44 pour 3 supports à 112,38 pour 4 supports et à 325,0 pour 5 supports. Ces résultats sont de 3 à 5 fois plus mauvais que ceux du modèle de Kwerel.

On peut donc conclure que l'hypothèse de base du modèle, selon laquelle l'interdépendance des audiences n'exclut pas l'indépendance des non couvertures, n'est pas vérifiée par les données.

25. Le Modèle Goodhart-Erhemberg

Ce modèle a été développé à partir des audiences des programmes télévisés en Angleterre. Il est fondé sur l'observation d'une relation quasiment constante liant les duplications observées avec les duplications qui résulteraient du pur hasard. Selon Goodhart-Erhemberg (24), le rapport des duplications observées aux duplications aléatoires serait constant quelque soit la paire de supports

$$(34) \qquad P_{ij} / (P_i * P_j) = K \qquad \forall_{i,j} \in M, \; i \neq j$$

Si l'on admet cette hypothèse, K peut être estimé en calculant la moyenne des K_{ij} observée pour chaque paire de supports.

$$(35) \qquad \hat{K} = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} K_{ij} / \binom{m}{2} \qquad \text{avec } K_{ij} = P_{ij} / (P_i * P_j)$$

Afin de tester l'hypothèse de Goodhart et Erhemberg dans le domaine des supports imprimés, nous avons calculé la valeur de K_{ij} pour chacune des 55 paires de magazines de l'enquête Politz de 1964. Nous avons ensuite calculé la valeur moyenne de K pour les 55 paires et, à l'aide de cette valeur moyenne, nous avons enfin calculé la duplication estimée pour chacune des paires de supports.

$$(36) \quad \hat{A}_{ij} = \hat{K} \left[(A_i * A_j) / POP \right]$$

Il ne nous reste plus qu'à comparer la distribution des duplications estimées, obtenues par la formule (36) avec la distribution des duplications réelles pour les 55 paires de supports. La qualité de l'ajustement des deux distributions sera mesurée par le test d'ajustement du Chi Deux, avec 54 degrés de liberté seulement, puisque $\hat{K}$ a été estimé à partir des données. Les résultats de ce test, pour six segments de population, apparaissent dans la table suivante :

Test de l'hypothèse de Duplication Constante de Goodhart-Erhemberg pour un échantillon de 11 magazines

Segment de Population	K Moyen	K Maximum	K Minimum	Chi Deux	Nombre d'individus enquêtés
Age 18 ans et +					
Ensemble	1,971	3,255	1,221	1 661,4	6 893
Femmes	1,665	2,405	1,214	417,9	4 412
Hommes	2,429	4,625	1,181	761,8	2 481
Age 18 à 34	1,694	3,147	0,922	601,1	2 261
Age 35 à 49	1,996	3,850	1,029	495,5	1 708
Age 50 et +	2,382	5,255	1,245	757,9	2 292

Pour tous les segments, on est conduit à rejeter l'hypothèse de la constance de K, au seuil de confiance de 99%. La duplication des supports imprimés ne se conforme donc pas au schéma de duplication constante observé pour les programmes télévisés.

26. Les Modèles d'Agrégation Séquentielle

Goodhart et Erhemberg n'indiquent pas comment estimer la couverture nette de combinaisons comportant plus de deux supports. Il est cependant aisé d'étendre leurs formules à m supports en recourant au processus d'agrégation séquentielle qui consiste à estimer tout d'abord la couverture nette de 2 supports, puis la couverture nette du nouveau support agrégé avec un 3ième support et ainsi de suite jusqu'à épuisement des supports. Par exemple, si nous

considérons 3 supports 1, 2, 3, la couverture nette des deux premiers sera égale à :

$$R_{1,2} = A_1 + A_2 - K * \left[A_1 * A_2/POP\right]$$

La couverture nette des 3 supports s'obtient en combinant $R_{1,2}$ avec le 3ième support :

$$R_{12,3} = R_{1,2} + A_3 - K * \left[R_{12} * A_3/POP\right]$$

(37)
$$R_{12,3} = A_1 + A_2 + A_3 - \frac{K}{POP}\left[A_1*A_2+A_1*A_3+A_2*A_3\right] + \left(\frac{K}{POP}\right)^2 A_1*A_2*A_3$$

Nous observons que, grâce à la constance supposée de K, la couverture nette des 3 supports ne dépend pas de l'ordre d'inclusion de ceux-ci.

$$R_{12,3} = R_{13,2} = R_{23,1} = R_{123}$$

La formule (37) peut être réécrite de la façon suivante :

$$R_{123} = \Sigma_{i=1}^{3} A_i - K\,\Sigma_{i=1}^{2}\,\Sigma_{j=i+1}^{3}(A_i*A_j/POP) + K^2(A_i*A_j*A_k/POP^2)$$

Les quantités $A_i \times A_j/POP$ et $A_i * A_j * A_k/POP^2$ ne sont autres que les audiences dupliquées et tripliquées dans l'hypothèse d'indépendance des audiences. Nous obtenons donc :

$$R_{123} = S_1 - K\,S_2 + K^2 S_3$$

Cette dernière expression peut être aisément généralisée à m supports

(38) $$R_m = \Sigma_{i=1}^{m} (-1)^{i-1} K^{i-1} S_i$$

avec S_i = Somme des recouvrements d'ordre i, dans l'hypothèse d'indépendance des audiences des supports.

L'équation (38) est similaire à l'équation (4) qui donne la valeur exacte, par définition, de la couverture nette. L'hypothèse implicite du modèle d'agrégation séquentielle avec duplication constante peut donc s'exprimer comme suit :

Somme des recouvrements d'ordre i = $K^{i-i} S_i$

Nous pouvons démontrer maintenant que lorsque K est estimé par l'équation (35), la somme des duplications estimées par le modèle de duplication constante n'est pas égale à la somme des duplications observées. En d'autres termes nous allons montrer que $K * S_2$ est différent de D, à moins que tous les K_{ij} soient égaux.

$$(39) \quad K = POP \left[\Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} (A_{ij} / A_i * A_j) \right] / \binom{m}{2}$$

$$(40) \quad \frac{D}{S_2} = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij} / \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} (A_i * A_j / POP)$$

Les expressions (39) et (40) ne sont pas égales sauf si tous les rapports, $K_{ij} = A_{ij} / (A_i * A_j)$, sont égaux. Or, nous avons vu que l'hypothèse de l'égalité de tous les K_{ij} n'est pas vérifiée pour les magazines. Nous pouvons donc conclure qu'il est possible d'améliorer le modèle d'agrégation séquentielle en abandonnant l'hypothèse de duplication constante et en remplaçant K par la duplication variable K' :

$$(41) \quad K' = \frac{D}{S_2} = \frac{\text{Somme des duplications observées}}{\text{Somme des duplications aléatoires}}$$

Nous appellerons ce nouveau modèle agrégation séquentielle avec duplication variable. On peut remarquer que K' est équivalent à K lorsque 2 supports seulement sont considérés. Quelque soit le coefficient utilisé, la distribution des occasions de voir sera donnée par le théorème de Waring. En combinant l'expression (7) avec l'expression (38), nous obtenons :

$$(42) \quad E_k^m = \Sigma_{i=k}^{m} (-1)^{i-k} * \binom{i}{k} * K^{i-1} * S_i$$

La table suivante compare les résultats obtenus par les 2 modèles d'agrégation séquentielle, pour les données de Marc et pour les données de Politz.

Comparaison des modèles d'Agrégation Séquentielle à Duplication Constante et à Duplication Variable

Combinaison	Couverture Nette Observée	Couverture Nette estimée					
		Duplication Constante			Duplication Variable		
		K	Couverture	%Erreur	K'	Couverture	%Erreur
Données de Marc							
3 Magazines							
A-B-C	372	2,734	360	-3,22	2,727	361	-2,96
A-B-D	354	2,659	354	0,00	2,701	351	+ ,84
A-B-E	349	2,843	340	-2,58	2,839	341	-2,29
A-C-D	365	2,542	369	+1,09	2,571	366	+ ,27
A-C-E	356	2,861	341	-4,21	2,843	343	-3,65
A-D-E	329	2,666	330	+ ,30	2,683	328	- ,30
B-C-D	371	2,618	369	- ,54	2,628	368	-1,88
B-C-E	364	2,805	351	-3,57	2,784	353	-3,02
B-D-E	349	2,696	345	-1,15	2,707	344	-1,43
C-D-E	345	2,712	348	+ ,86	2,741	346	+ ,29
4 Magazines							
A-B-C-D	382	2,638	374	-2,09	2,662	371	-2,88
A-B-C-E	378	2,811	354	-6,35	2,792	356	-5,82
A-B-D-E	366	2,716	359	-1,91	2,740	356	-2,73
A-C-D-E	370	2,695	363	-1,89	2,714	361	-2,43
B-C-D-E	374	2,708	365	-2,41	2,715	364	-2,67
5 Magazines							
A-B-C-D-E	384	2,713	367	-4,42	2,727	366	-4,69
Données de Politz							
Age 18 ans ou +	74 630	1,519	70 508	-5,52	1,510	70 770	-5,17
Femmes 18 ou +	40 710	1,533	37 682	-7,44	1,511	38 072	-6,48
Hommes 18 ou +	33 920	1,528	32 510	-4,16	1,518	32 637	-3,78
11 Magazines							
Age 18 ans ou +	83 180	1,971	59 650	-28,3	1,766	66 578	-19,9
Femmes 18 ou +	47 300	1,665	35 738	-24,4	1,636	36 589	-22,6
Hommes 18 ou +	35 880	2,429	23 870	-33,5	1,698	31 781	-11,4
Age 18 à 34	30 800	1,694	23 286	-24,3	1,505	26 053	-15,4
Age 35 à 49	25 610	1,996	17 502	-31,7	1,816	19 407	-24,2
Age 50 ans ou +	26 770	2,382	18 134	-32,3	2,085	20 702	-22,7

Commentaires

On observe que le coefficient de duplication constante K, calculé par la formule (39), est très similaire au coefficient de duplication variable K', calculé par la formule (40), du moins lorsque le nombre de supports n'est pas trop élevé. Par contre, pour la combinaison de 11 magazines, le modèle de duplication variable s'avère nettement supérieur au modèle de duplication constante. Quelle que soit la valeur de K utilisé, on observe une sous-estimation systématique de la couverture nette. Cette sous-estimation s'accroît en fonction du nombre de supports, comme le montre la table suivante.

Pourcentage Moyen de Sous-estimation des Modèles d'Agrégation Séquentielle

	Données de Marc			Données de Politz	
Nombre de Supports	3 supports	4 supports	5 supports	5 supports	11 Supports
Duplication constante	1,75%	2,93%	4,42%	5,70%	29,08%
Duplication variable	1,69%	3,30%	4,69%	5,10%	19,38%

Ces résultats sont similaires à ceux obtenus par les modèles de Kwerel ou de Sainsbury modifié pour les données de Marc. Ils sont nettement moins bons en ce qui concerne la combinaison de 11 supports puisque le pourcentage moyen d'erreurs du meilleur modèle d'agrégation séquentielle est de 19,4% contre 13,5% pour le modèle de Kwerel et 12,7% pour le modèle de Sainsbury modifié.

Les résultats de l'estimation de la distribution des occasions de voir sont présentés en détail dans les tables de la page 397 à la page 401 . On observe un accroissement modéré de la valeur du Chi Deux moyen qui est pratiquement identique pour les deux modèles d'agrégation séquentielle. Le Chi Deux moyen est égal à 8,8 pour 3 supports, 18,4 pour 4 supports et 29,4 pour 5 supports. Ces résultats sont les meilleurs obtenus jusqu'à présent. Dans le cas de 5 supports, le Chi Deux est 11 fois plus faible que celui du modèle de Sainsbury modifié et 3 fois plus faible que celui du modèle de Kwerel.

27. Le Modèle d'Indépendance Conditionnelle

Ce modèle, comme les précédents, est basé sur l'hypothèse que les duplications ne sont pas aléatoires. En d'autres termes, les probabilités inconditionnelles d'exposition ne sont pas indépendantes.

$$(43) \qquad P_{ij} \neq P_i * P_j \qquad \forall_{i,j} \in M \ ; \ i \neq j$$

Cependant, cette hypothèse n'exclut pas la possibilité que les probabilités conditionnelles d'exposition puissent être indépendantes par paires.

$$(44) \qquad P_{ij/k} = P_{i/k} * P_{j/k} \qquad \forall_{i,j} \in M \ ; \ i \neq j \neq k$$

Le modèle procède tout d'abord à l'estimation des triplications à partir des duplications, puis à l'estimation des quadruplications à partir des triplications, et ainsi de suite. Par définition des probabilités conditionnelles, nous avons :

$$P_{i/k} = P_{ik} / P_k \quad ; \quad P_{ij/k} = P_{ijk}/P_k \quad ; \quad \ldots \text{ etc}$$

En combinant ces équations avec l'hypothèse (44), nous obtenons la valeur estimée des triplications.

$$(45) \qquad P_{ijk} = P_{ik} * P_{jk}/P_k = P_{ij} * P_{jk}/P_j = P_{ij} * P_{ik}/P_i$$

En termes d'audiences, l'expression (45) devient :

$$A_{ijk} = A_{ik} * A_{jk}/A_k = A_{ij} * A_{jk}/A_j = A_{ij} * A_{ik}/A_i$$

Nous pouvons donc estimer les triplications en prenant la moyenne des 3 estimations possibles.

$$(46) \qquad \hat{A}_{ijk} = \frac{1}{3}\left[(A_{ik} + A_{jk}/A_k) + (A_{ij} * A_{jk}/A_j) + (A_{ij} * A_{ik}/A_i)\right]$$

Finalement, la couverture nette et la distribution des occasions de voir peut être estimée à partir des sommes d'audiences, des sommes de duplications et des sommes estimées de recouvrement d'ordre supérieur à deux, en appliquant le théorème de Waring. L'exemple numérique suivant présente le détail des calculs pour une combinaison de 4 magazines extraite des données de Marc.

Exemple Numérique

Magazines	Audiences	Paires	Duplications	Paires	Duplications
A	225	AB	183	BC	228
B	285	AC	181	BD	117
C	300	AD	84	CD	115
D	153				

Somme des Audiences = 963 Somme des Duplications = 908

Nous estimons tout d'abord les triplications par l'équation (46) :

$$\left.\begin{aligned}
\widehat{ABC} &= \frac{1}{3}\ (147{,}2 + 137{,}6 + 146{,}4) = 143{,}7 \\
\widehat{ABD} &= \frac{1}{3}\ (68{,}3 + 64{,}2 + 75{,}1) = 69{,}2 \\
\widehat{BCD} &= \frac{1}{3}\ (93{,}6 + 87{,}4 + 87{,}9) = 89{,}6 \\
\widehat{ACD} &= \frac{1}{3}\ (67{,}6 + 63{,}1 + 69{,}4) = 66{,}7
\end{aligned}\right\} \text{Somme des triplications} = 369$$

Nous pouvons maintenant estimer la quadruplication :

$$\widehat{ABCD} = \frac{1}{6}\ (143{,}7*69{,}2/183 + 143{,}7*89{,}6/228 + 89{,}6*66{,}7/115 + 69{,}2*89{,}6/117 + 66{,}7*143{,}7/181) = 54{,}7$$

La couverture nette estimée s'obtient en appliquant l'expression (4) :

$$R_{ABCD} = 963 - 908 + 369 - 55 = 369$$

La couverture nette mesurée par Marc étant de 382, le modèle d'indépendance conditionnelle conduit, dans ce cas, à une sous-estimation de 3,4%.

La distribution estimée des occasions de voir s'obtient à l'aide du théorème de Waring :

Nombre d'expositions		Distribution Estimée	Distribution Observée
0 sur 4	E_0^4	= 100-963+908-369+55 = 631	618
1 sur 4	E_1^4	= 963-2*908+3*369-4*55 = 34	66
2 sur 4	E_2^4	= 908-3*369+4*55 = 131	113
3 sur 4	E_3^4	= 369-4*55 = 149	141
4 sur 4	E_4^4	= 55 = 55	62

La valeur du test d'ajustement du Chi Deux est de 19,9. Les résultats détaillés du modèle d'indépendance conditionnelle pour les 14 autres combinaisons de supports des données de Marc sont présentées dans les tables de la page 397 à la page 401 . On observe que le modèle sous-estime systématiquement la couverture nette. Néanmoins, le pourcentage d'erreur n'est pas plus mauvais que ceux observés pour les modèles d'agrégation séquentielle et pour le modèle de Kwerel. Les résultats sont même nettement meilleurs que ceux obtenus par le modèle de Sainsbury modifié, sans toutefois être aussi bon que ceux obtenus par le modèle d'Hofmans.

En ce qui concerne l'estimation de la distribution des occasions de voir, on observe à nouveau des résultats très comparables à ceux des modèles d'agrégation séquentielle et légèrement inférieurs à ceux du modèle de Kwerel. Toutefois, le principal inconvénient du modèle d'indépendance conditionnelle tient à la lourdeur des calculs. Si la combinaison comprend m supports, le nombre total d'opérations qui devront être effectuées est égal à :

$$\Sigma_{i=3}^{m} \binom{m}{i} \binom{i}{i-2}$$

Par exemple, dans le cas de 11 supports, cela représente 56 219 opérations !

28. Le Modèle D'Hofmans Etendu

Nous avons déjà étudié la méthode d'Hofmans lors de notre présentation des modèles de duplication du type Ad-Hoc. Toutefois, la mise à jour de l'hypothèse implicite concernant les recouvrements d'ordre supérieur à deux, permet de considérer qu'il s'agit d'un modèle stochastique caractérisé par les trois hypothèses suivantes :

1 - Les individus sont homogènes

2 - Les supports sont hétérogènes

3 - Les audiences sont interdépendantes. La forme de cette interdépendance est définie par l'hypothèse implicite du modèle d'Hofmans.

$$D^{n-1}/A^{n-2} \leq \text{Somme des recouvrements d'ordre } n \leq (2D)^{n-1}/A^{n-2}$$

Le principal inconvénient de la formule d'Hofmans est qu'elle ne permet pas d'estimer la distribution des occasions de voir. Son principal avantage est de conduire, pour nos données, aux meilleures estimations de la couverture nette. De plus, elle ne nécessite pas de calculs volumineux et enfin le pourcentage d'erreur est relativement peu sensible au nombre de supports inclus. En conséquence, nous avons développé un nouveau modèle, baptisé Hofmans étendu, qui tout en conservant les avantages de la formule d'Hofmans, permet également d'estimer la distribution des occasions de voir.

Pour cela, nous allons démontrer tout d'abord que la somme des couvertures nettes de toutes les combinaisons de supports parmi m peut s'exprimer en fonction des sommes de recouvrement d'ordre un jusqu'à l'ordre v. Nous utiliserons la notation suivante :

$\Sigma R_{v/m}$ = Somme des couverture nettes de toutes les combinaisons de v supports parmi m

S_v = Somme des recouvrements d'ordre v.

Nous établirons tout d'abord la relation pour v = 1,2,3, puis nous généraliserons par induction.

- Lorsque v=1, la solution est triviale puisque la couverture nette d'un support est égale à son audience et qu'il n'existe que m façons de choisir un support parmi m.

$$\Sigma R_{1/m} = \Sigma_{i=1}^{m} A_i = S_1$$

- Lorsque v = 2, la couverture nette de chacune des $\binom{m}{2}$ combinaisons de deux supports est donnée par : $R_2 = A_i+A_j-A_{ij}$. Si nous considérons les audiences de chaque support, nous en comptons 2 pour chacune des $\binom{m}{2}$ combinaisons possibles. Comme nous avons m audiences différentes, chacune d'elle sera donc comptée $2*\binom{m}{2}/m$ fois, c'est à dire $\binom{m-1}{1}$ fois. Si nous considérons les duplications, nous en comptons une pour chacune des $\binom{m}{2}$ combinaisons de supports, et comme il y en a $\binom{m}{2}$ au total, chacune d'elle n'est comptée qu'une fois. En conséquence, la somme de toutes les couvertures nettes obtenues en choisissant 2 supports parmi m est égale à :

$$\Sigma R_{2/m} = \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_i+A_j-A_{ij} = \binom{m-1}{1} \Sigma_{i=1}^{m} A_i - \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij}$$

$$\Sigma R_{2/m} = \binom{m-1}{1} * S_1 - S_2$$

- Lorsque v = 3, la couverture nette de chacune des $\binom{m}{3}$ combinaisons de 3 supports parmi m est donnée par : $R_3 = A_i+A_j+A_k-A_{ij}-A_{ik}-A_{jk}+A_{ijk}$
Si nous considérons les audiences, nous en comptons 3 pour chacune des $\binom{m}{3}$ combinaisons possibles. Comme nous n'avons que m supports, chaque audience aura donc été comptée $3 * \binom{m}{3}/m$ fois, c'est à dire un total de $\binom{m-2}{2}$ fois. Si nous considérons les duplications, nous en comptons 3 pour chaque combinaison, c'est à dire au total $3 * \binom{m}{3}$ alors qu'il n'en existe que $\binom{m}{2}$. Chaque duplication a donc été comptée $\binom{m-2}{1}$ fois. Les triplications sont comptées une fois pour chacune des $\binom{m}{3}$ combinaisons et comme il y a $\binom{m}{3}$ triplications, chacune d'elle n'est comptée qu'une fois. Nous obtenons donc :

$$R_{3/m} = \binom{m-1}{2} S_1 - \binom{m-2}{1} S_2 + S_3$$

Sans qu'il soit nécessaire d'aller plus loin nous pouvons généraliser les résultats obtenus au cas de combinaisons de v supports parmi m. Nous obtenons :

$$(47) \quad \Sigma R_{v/m} = \binom{m-1}{v-1} * S_1 - \binom{m-2}{v-2} * S_2 + \binom{m-3}{v-3} * S_3 - \ldots\ldots \pm S_v$$

Nous pouvons réécrire cette expression sous une forme plus compacte :

$$(48) \quad \Sigma R_{v/m} = \Sigma_{i=1}^{v} (-1)^{i+1} * \binom{m-i}{v-i} * S_i$$

Cette équation va nous permettre d'estimer les sommes inconnues S_v en fonction des $\Sigma R_{v/m}$ qui sont aisément calculées à l'aide de la formule d'Hofmans. Il nous suffit pour cela de résoudre l'équation (48) pour les valeurs de S_v. La solution est donnée par les deux équations suivantes, pour les valeurs paires et impaires de v.

$$\text{(49)} \quad \begin{cases} \text{Si v est pair : } S_v = \Sigma_{i=1}^{v-1} (-1)^{i+1} * \binom{m-i}{v-i} * S_i - \Sigma R_{v/m} \\ \text{Si v est impair : } S_v = \Sigma_{i=1}^{v-1} (-1)^{i} * \binom{m-i}{v-i} * S_i + \Sigma R_{v/m} \end{cases}$$

Finalement, à l'aide des valeurs estimées de S_v et du théorème de Waring, nous pouvons estimer la distrivution des occasions de voir

$$\text{(50)} \quad E_k^m = \Sigma_{i=1}^{m} (-1)^{i-k} \binom{i}{k} * S_i .$$

L'exemple numérique suivant, basé sur les données de Marcn décrit le détail des calculs pour une combinaison de 5 supports.

Exemple numérique

Supposons que nous désirions estimer la distribution des occasions de voir générée par les 5 magazines A,B,C,D,E, mesurés par Marc. En appliquant la formule d'Hofmans, nous estimons tout d'abord les couvertures nettes de toutes les combinaisons de 3,4 et 5 supports, puis nous calculons les sommes de couvertures nettes estimées pour 3,4 et 5 supports.

$\Sigma R_{3/5}$= 376+361+354+378+357+337+382+369+356+357 = 3 627

La somme observée est de 3 554. Le pourcentage d'erreur est de 2,1%

$\Sigma R_{4/5}$ = 400+386+382+388+393 = 1 949

La somme observée est de 1 870. Le pourcentage d'erreurs est de 4,2%

$\Sigma R_{5/5} = R_5$ = 407. La couverture nette réelle est de 384, soit un pourcentage d'erreurs de 6%. En utilisant la formule (49), nous calculons maintenant les S_v estimés :

S_1 = 1 146 $\quad \hat{S}_3$ = 981 $\quad \hat{S}_5$ = 57

S_2 = 1 410 $\quad \hat{S}_4$ = 367

En utilisant le théorème de Waring, nous obtenons la distribution estimée des occasions de voir.

Expositions	0	1	2	3	4	5
Observée	616	50	86	108	100	40
Estimée	593	88	93	92	76	59

Les tables de la page 397 à la page 401 présentent la distribution estimée des occasions de voir pour toutes les combinaisons de supports incluses dans les données de Marc. Nous observons que le Chi Deux moyen n'est que de 9,9 pour les 10 combinaisons de 3 supports. Ce résultat est nettement supérieur à celui du modèle de Sainsbury modifié. Il est équivalent à celui des modèles d'agrégation séquentielle mais nettement moins bon que le résultat obtenu par le modèle de Kwerel. Pour les 5 combinaisons de 4 supports, le Chi Deux moyen du modèle d'Hofmans étendu est de 26,4 contre 24,7 pour le modèle de Kwerel. Pour la combinaison de 5 supports, le modèle d'Hofmans étendu procure un ajustement de meilleure qualité que le modèle de Kwerel.

Finalement, il convient de remarquer que la formule combinatoire que nous avons établi, et qui lie les sommes de couvertures nettes aux sommes de recouvrement, est une identité mathématique et non pas une formule d'approximation. En conséquence, la distribution des occasions de voir serait exactement estimée si les couvertures nettes pouvaient être connues sans erreur. Par contre, lorsque des erreurs d'estimation se produisent au niveau de la couverture nette, elles sont amplifiées par les formules (49) et (50), et la qualité de l'estimation de la distribution des occasions de voir se détériore rapidement.

Nous avons terminé notre analyse des modèles qui assument une population cible homogène. Nous allons considérer maintenant un modèle de duplication qui permet de tenir compte de la variation des probabilités d'exposition d'un individu à un autre.

29. Le Modèle de Metheringham (25)

Hypothèses :

1 - Les individus sont hétérogènes
2 - Les supports sont homogènes
3 - Les audiences sont interdépendantes

Ce modèle, à la différence des précédents, conçoit les individus comme possédant chacun leur propre probabilité d'exposition à un support donné. Cette probabilité caractérise leur degré de fidélité par rapport au support et on suppose de plus que cette probabilité demeure stationnaire dans le temps. Enfin, bien que Metheringham ne le mentionne pas explicitement, son modèle revient à supposer que les probabilités individuelles d'exposition suivent une loi Bêta, quelle que soit la population cible considérée. Metheringham assume aussi l'homogénéité des supports et remplace les audiences et les duplications de chaque paire de supports par leur moyenne. Ce faisant, il considère qu'une insertion dans chacun des m supports est équivalente à m parutions d'un hypothétique support moyen. Chaque parution constituant un tirage indépendant, l'exposition à k parutions parmi m suit alors une loi de probabilité Bêta-Binomiale dont il s'agit d'estimer les paramètres en fonction de la moyenne des audiences et de la moyenne des duplications.

Ceci peut être accompli de trois façons différentes que nous allons présenter en détail. L'approche directe, basée sur l'audience moyenne et la duplication moyenne, n'est pas utilisée par Metheringham. Sa solution est fondée sur l'utilisation du concept de non couverture. Nous noterons NC_i la non couverture du support i :

$$(51) \qquad NC_i = POP - A_i \qquad \forall_i \in M$$

De même, la non couverture d'une paire de supports est définie de la façon suivante :

$$(52) \qquad NC_i = POP - R_{ij} = POP - A_i - A_j + A_{ij}$$

A partir des données d'audience, Metheringham calcule la non couverture moyenne des m supports, $\overline{NC}$, et la non couverture moyenne des $\binom{m}{2}$ paires de supports, $\overline{NC}_{ij}$

$$(53) \qquad \overline{NC} = \frac{1}{m} \Sigma_{i=1}^{m} NC_i = POP - \overline{A}$$

$$(54) \qquad \overline{NC_{ij}} = \frac{1}{\binom{m}{2}} \Sigma_{i-1}^{m-1} \Sigma_{j=i+1}^{m} NC_{ij} = POP - 2\,\overline{A} + \overline{D}$$

La non couverture moyenne de m supports, $\overline{NC}$, est équivalente au nombre d'individus non exposés à une parution, E_0^1, du support moyen hypothétique. Par application de la loi Bêta Binomiale, nous obtenons donc :

$$(55) \qquad \overline{NC} = E_0^1 = \left(\frac{b}{a+b}\right) * POP$$

De même, la non couverture moyenne des $\binom{m}{2}$ paires de supports est équivalente au nombre d'individus non exposés à deux parutions, E_0^2, du support moyen hypothétique.

$$(56) \qquad \overline{NC_{ij}} = E_0^2 = \left[\frac{(b+1)\ b}{(a+b+1)(a+b)}\right] * POP$$

En conséquence, nous pouvons estimer les paramètres a et b de la distribution Bêta des probabilités individuelles en résolvant les équations (55) et (56).

Nous posons $q = \overline{NC}/POP$ et $q_{ij} = \overline{NC}_{ij}/POP$, et nous obtenons :

$$(57) \qquad \hat{b} = \frac{q^2 - q * q_{ij}}{q_{ij} - q^2} \qquad \hat{a} = \hat{b} * (1-q)/q$$

La solution de Metheringham, quoique élégante, n'est guère intuitive. Il paraît plus simple d'estimer directement les paramètres a et b à partir de l'audience moyenne $\bar{A}$ et de la duplication moyenne $\bar{D}$, sans passer par le concept de non couverture. L'audience moyenne de la combinaison de m supports n'est autre que l'audience d'une parution, E_1^1, du support moyen hypothétique.

$$(58) \qquad \bar{A} = E_1^1 = \left[\frac{a}{a+b}\right] * POP$$

De même, la duplication moyenne $\bar{D}$ n'est autre que le nombre d'individus exposés à exactement 2 parutions du support moyen hypothétique :

$$(59) \qquad \bar{D} = E_2^2 = \left[\frac{(a+1)\ a}{(a+b+1)(a+b)}\right] * POP$$

En posant $p = \bar{A}/POP$ et $P_{ij} = \bar{D}/POP$, nous obtenons :

$$(60) \qquad \hat{a} = \frac{p^2 - p * p_{ij}}{p_{ij} - p^2} \qquad \hat{b} = a * (1-p)/p$$

Bien entendu, les estimations des équations (60) et (57) sont égales, comme on peut le vérifier en utilisant les équations (53) et (54)

Les deux approches que nous venons de présenter ne sont pas totalement satisfaisantes. Elles masquent le fait que les paramètres a et b sont estimés, par la méthode des moments, à partir de l'information contenue dans la somme des audiences et dans la somme des duplications sur la moyenne et la variance de la distribution des occasions de voir. Pour le démontrer, il suffit de se rappeler qua le moyenne et la variance de la distribution Bêta Binomiale sont égales à :

$$(61) \quad \begin{cases} \mu = m * a / (a+b) \\ \sigma^2 = m * a * b * (m+a+b) / \left[(a+b)^2(a+b+1)\right] \end{cases}$$

Il s'en suit que les paramètres a et b peuvent être exprimés en fonction de la moyenne et de la variance :

$$(62) \quad \begin{cases} a = \mu\left[\sigma^2-(m-\mu)\right] / \left[\mu(m-\mu)-m\,\sigma^2\right] \\ b = a * (m-\mu) / \mu \end{cases}$$

Par ailleurs, la moyenne et la variance de la distribution des occasions de voir peuvent être exprimées en fonction de la somme des audiences et de la somme des duplications, ainsi que nous l'avons montré au début de ce chapitre :

$$(63) \quad \begin{cases} \mu = A / POP \\ \sigma^2 = (A+2D) / POP - (A/POP)^2 \end{cases} \qquad \begin{aligned} A &= \Sigma_{i=1}^{m} A_i \\ D &= \Sigma_{i=1}^{m-1} \Sigma_{j=i+1}^{m} A_{ij} \end{aligned}$$

En remplaçant dans l'équation (62) μ et σ^2 par leurs valeurs données par l'équation (63), nous obtenons l'estimation par la méthode des moments des paramètres a et b.

$$(64) \quad \begin{cases} \hat{a} = \left[2*D*A - A^2*(m-1)\right] / \left[A^2*(m-1) - 2*D*m*POP\right] \\ \hat{b} = \hat{a} * (m * POP - A)/A \end{cases}$$

Bien entendu, les estimations de l'équation (64) sont identiques à celles des équations (60) et (57). Pour le démontrer, il suffit de remplacer dans l'équation (60) $\overline{A}$ par A/m et $\overline{D}$ par 2D/m(m-1), ou bien de remplacer dans l'équation (57) $\overline{NC}$ par POP - A/m et $\overline{NC_{ij}}$ par POP - 2A/m(m-1).

L'équation (64) montre que la formule de Metheringham n'est rien d'autre que l'estimation par la méthode des moments des paramètres de la distribution Bêta Binomiale s'ajustant le mieux à la distribution des occasions

de voir. Metheringham (25) a donc tort lorsqu'il énonce que sa formule est totalement empirique et qu'elle n'a aucun lien avec la distribution Bêta Binomiale.

Les paramètres a et b étant estimés, la distribution des occasions de voir s'obtient par la loi Bêta Binomiale :

$$(65) \qquad E_k^m = POP * \binom{m}{k} * \frac{(a+k-1) \ldots (a)\ (b+m-K-1) \ldots (b)}{(a+b+m-1) \ \ldots\ldots\ldots\ldots\ (a+b)}$$

La couverture nette s'en déduit aussitôt en soustrayant à la population totale le nombre d'individus non exposés :

$$(66) \qquad R_m = POP - E_o^m$$

L'exemple numérique suivant présente le détail des calculs pour une combinaison de 4 supports extraite des données de Marc.

Exemple Numérique

Magazines	Audiences	Paires	Duplications	Paires	Duplications
A	225	AB	183	BC	228
B	285	AC	181	BD	117
C	300	AD	84	CD	115
D	115				
		Somme des duplications = 908			
Somme =	963	Population = 1000 ; m = 4			

Nous calculons tout d'abord l'audience moyenne et la duplication moyenne, puis les probabilités p et p_{ij}

$\bar{A} = 963 / 4 = 241$ $\qquad$ $\bar{D} = 908 / 6 = 151$

$p = 241/1000 = 0,241$ $\qquad$ $P_{ij} = 151 / 1000 = 0,151$

En utilisant l'équation (60) nous estimons les paramètres a et b :

$\hat{a} = (0,241^2 - 0,241 * 0,151) / (0,151 - 0,241^2) = 0,231$

$\hat{b} = 0,231\ (1 - 0,241) / 0,241 = 0,727$

Nous pouvons maintenant estimer la distribution des occasions de voir et la comparer à la distribution observée :

Distribution des Occasions de Voir		Estimée	Observée
$E_0^4 = \frac{(0,727)(1,727)(2,727)(3,727)}{(0,958)(1,958)(2,958)(3,958)} * 1\ 000$	=	582	618
$E_1^4 = \frac{(4)(0,231)(0,727)(1,727)(2,727)}{(0,958)(1,958)(2,958)(3,958)} * 1\ 000$	=	144	66
$E_2^4 = \frac{(6)(0,231)(1,231)(0,727)(1,727)}{(0,958)(1,958)(2,958)(3,958)} * 1\ 000$	=	97	113
$E_3^4 = \frac{(4)(0,231)(1,231)(2,231)(0,727)}{(0,958)(1,958)(2,958)(3,958)} * 1\ 000$	=	84	141
$E_4^4 = \frac{(0,231)(1,231)(2,231)(3,231)}{(0,958)(1,958)(2,958)(3,958)} * 1\ 000$	=	93	62

Nous avons testé le modèle de Metheringham pour les données de Politz et pour les données de Marc. La table suivante présente les résultats du modèle de Metheringham, en ce qui concerne l'estimation de la couverture nette, pour les données de Politz.

Estimation de la Couverture Nette pour le Modèle de Metheringham

Segment de Population	Couverture Nette		Pourcentage d'erreurs
	Observée	Estimée	
Combinaison de 5 magazines, Age 18 ans ou +			
Ensemble	74 630	75 367	+ 0,99
Femmes	40 710	40 851	+ 0,35
Hommes	33 920	34 599	+ 2,00
Combinaison de 11 magazines, Age 18 ans ou +			
Ensemble	83 180	82 564	- 0,74
Femmes	47 300	47 624	+ 0,68
Hommes	35 880	36 514	+ 1,77
Age 18 à 34 ans	30 800	30 957	+ 0,51
Age 35 à 49 ans	25 610	24 697	- 3,56
Age 50 ans ou +	26 770	26 375	- 1,47

Cette table fait apparaître l'excellente qualité des estimations du modèle de Metheringham. Les tables suivantes comparent, sur la base des données de Marc, la distribution des occasions de voir estimée par 7 modèles de duplications.

Test d'Ajustement de la Distribution des Occasions de Voir sur la base des données de Marc

Combinaisons de 3 supports	Distribution observée	Kwerel EST	Kwerel %E	Sainsbury EST	Sainsbury %E	Agré. Var. EST	Agré. Var. %E	Agré. Const. EST.	Agré. Const. %E	Indé. Cond. EST	Indé. Cond. %E	Hofmans EST.	Hofmans %E	Metheringham EST	Metheringham %E
A-B-C															
READ NONE	628	624	.64	605	3.66	639	-1.75	640	-1.91	638	-1.59	624	.64	617	1.75
READ ONE	88	100	-13.64	156	-77.27	55	37.50	54	38.64	57	35.23	100	-13.64	120	-36.36
READ TWO	130	119	8.46	62	52.31	163	-25.38	162	-24.62	161	-23.85	118	9.23	98	24.62
READ THREE	154	158	-2.60	177	-14.94	143	7.14	144	6.49	144	6.49	158	-2.60	165	-7.14
CHI SQUARE			2.6965		92.3921		21.7303		21.8919		19.1213		2.8734		20.4917
STD. PREDICTION ERROR			8.6168		50.7592		24.5967		24.6171		23.0326		8.9443		23.9270
A-B-D															
READ NONE	646	645	.15	639	1.08	649	-.46	646	0.00	652	-.93	639	1.08	630	2.48
READ ONE	120	122	-1.67	141	-17.50	110	8.33	115	4.17	103	14.17	142	-18.33	168	-40.00
READ TWO	159	157	1.26	138	13.21	169	-6.29	170	-6.92	176	-10.69	137	13.84	111	30.19
READ THREE	75	76	-1.33	82	-9.33	72	4.00	69	8.00	69	8.00	82	-9.33	91	-21.33
CHI SQUARE			.0734		7.1778		1.5962		1.4493		4.7617		7.8065		37.5002
STD. PREDICTION ERROR			1.5811		15.6525		7.3824		6.7454		12.7475		16.3248		35.7771
A-D-E															
READ NONE	651	651	0.00	636	2.30	659	-1.23	660	-1.38	659	-1.23	646	.77	639	1.84
READ ONE	108	109	-.93	152	-40.74	83	23.15	82	24.07	83	23.15	123	-13.89	143	-32.41
READ TWO	138	137	.72	94	31.88	163	-18.12	163	-18.12	163	-18.12	123	10.87	103	25.36
READ THREE	103	103	0.00	118	-14.56	95	7.77	95	7.77	95	7.77	108	-4.85	115	-11.65
CHI SQUARE			.0165		34.4850		11.0357		11.5340		11.0357		3.9949		21.8387
STD. PREDICTION ERROR			.7071		32.8710		18.5607		19.0132		18.5607		11.1803		26.1630
A-C-D															
READ NONE	635	632	.47	620	2.36	634	.15	631	.63	635	0.00	622	2.05	614	3.31
READ ONE	119	129	-8.40	163	-36.97	123	-3.36	127	-6.72	118	.84	158	-32.77	181	-52.10
READ TWO	179	169	5.59	135	24.58	175	2.23	176	1.68	180	-.56	140	21.79	117	34.64
READ THREE	67	70	-4.48	82	-22.39	68	-1.49	67	0.00	67	0.00	80	-19.40	88	-31.34
CHI SQUARE			1.5475		30.7971		.2403		.6133		.0140		24.0672		61.0540
STD. PREDICTION ERROR			7.3824		32.8710		2.9155		4.7170		.7071		29.0689		46.2871
A-C-E															
READ NONE	644	647	-.47	638	.93	657	-2.02	659	-2.33	656	-1.86	643	.16	636	1.24
READ ONE	117	108	7.69	134	-14.53	77	34.19	75	35.90	81	30.77	121	-3.42	140	-19.66
READ TWO	126	135	-7.14	109	13.49	166	-31.75	165	-30.95	162	-28.57	122	3.17	103	18.25
READ THREE	113	110	2.65	119	-5.31	100	11.50	101	10.62	101	10.62	114	-.88	121	-7.08
CHI SQUARE			1.4268		5.1382		28.1316		28.7721		22.8606		.2741		9.3855
STD. PREDICTION ERROR			6.7082		12.7475		29.7405		30.2242		26.8328		2.9155		17.2192

Test d'Ajustement de la Distribution des Occasions de Voir sur la base des données de Marc

Combinaison de 3 supports	Distribution Observée	Kwerel EST	Kwerel %E	Sainsbury EST	Sainsbury %E	Agré. Var. EST.	Agré. Var. %E	Agré. Con. EST.	Agré. Con. %E	Indé. Cond. EST.	Indé. Cond. %E	Hofmans EST	Hofmans %E	Metheringham EST.	Metheringham %E
A-D-E															
READ NONE	671	677	-.89	649	3.28	672	-.15	670	.15	672	-.15	663	1.19	658	1.94
READ ONE	143	124	13.29	210	-46.85	141	1.40	143	0.00	139	2.80	167	-16.78	181	-26.57
READ TWO	140	159	-13.57	73	47.86	142	-1.43	142	-1.43	144	-2.86	116	17.14	102	27.14
READ THREE	46	40	13.04	68	-47.83	45	2.17	45	2.17	45	2.17	54	-17.39	59	-28.26
CHI SQUARE			5.9393		74.6989		.0798		.0518		.2494		9.6289		24.3380
STD. PREDICTION ERROR			14.0890		49.8648		1.5811		1.2247		2.9155		17.8885		28.3989
B-C-D															
READ NONE	629	621	1.27	611	2.86	632	-.48	631	-.32	632	-.48	618	1.75	607	3.50
READ ONE	97	122	-25.77	152	-56.70	89	8.25	91	6.19	87	10.31	130	-34.02	163	-68.04
READ TWO	181	156	13.81	126	30.39	189	-4.42	189	-4.42	191	-5.52	148	18.23	115	36.46
READ THREE	93	101	-8.60	111	-19.35	90	3.23	90	3.23	90	3.23	104	-11.83	115	-23.66
CHI SQUARE			10.6863		51.8972		1.1245		.8279		1.6945		18.7368		74.9473
STD. PREDICTION ERROR			18.5607		40.9207		6.0415		5.3151		7.3824		24.5967		49.1935
B-C-E															
READ NONE	636	632	.63	626	1.57	647	-1.73	649	-2.04	645	-1.42	631	.79	623	2.04
READ ONE	92	104	-13.04	123	-33.70	60	34.78	57	38.04	65	29.35	107	-16.30	130	-41.30
READ TWO	140	128	8.57	109	22.14	172	-22.86	171	-22.14	167	-19.29	125	10.71	102	27.14
READ THREE	132	136	-3.03	142	-7.58	121	8.33	123	6.82	123	6.82	137	-3.79	145	-9.85
CHI SQUARE			2.7402		18.2247		19.5516		21.0589		13.8721		4.2815		27.5560
STD. PREDICTION ERROR			8.9443		23.0326		23.9270		24.6779		20.1246		11.1803		28.3989
B-D-E															
READ NONE	651	656	-.77	642	1.38	656	-.77	655	-.61	656	-.77	644	1.08	639	1.84
READ ONE	140	124	11.43	167	-19.29	126	10.00	128	8.57	125	10.71	161	-15.00	177	-26.43
READ TWO	146	162	-10.96	119	18.49	160	-9.59	160	-9.59	161	-10.27	125	14.38	109	25.34
READ THREE	63	58	7.94	72	-14.29	58	7.94	58	7.94	58	7.94	70	-11.11	75	-19.05
CHI SQUARE			4.0172		11.6104		3.1777		2.7924		3.5835		7.0236		21.6622
STD. PREDICTION ERROR			11.8533		20.1246		10.5119		9.7596		11.1803		15.6525		27.5045
C-D-E															
READ NONE	655	653	.31	644	1.68	654	.15	652	.46	655	0.00	643	1.83	636	2.90
READ ONE	116	123	-6.03	149	-28.45	119	-2.59	123	-6.03	115	.86	152	-31.03	174	-50.00
READ TWO	167	160	4.19	134	19.76	164	1.80	164	1.80	168	-.60	131	21.56	109	34.73
READ THREE	62	64	-3.23	73	-17.74	63	-1.61	62	0.00	62	0.00	74	-19.35	81	-30.65
CHI SQUARE			.7864		18.0452		.1491		.4900		.0146		21.4753		55.5174
STD. PREDICTION ERROR			5.1478		24.5967		2.2361		4.0927		.7071		26.8328		43.1567

A-B-C-D	bution observed	EST	%E	EST	%E	EST	%E	EST	%E	EST.	%E	EST.	%E	EST.	%E
READ NONE	618	599	3.07	575	6.96	629	-1.78	626	-1.29	631	-2.10	600	2.91	582	5.83
READ ONE	66	112	-69.70	176	-166.67	40	39.39	43	34.85	34	48.48	101	-53.03	144	-118.18
READ TWO	113	87	23.01	43	61.95	126	-11.50	129	-14.16	131	-15.93	113	0.00	97	14.16
READ THREE	141	130	7.80	125	11.35	150	-6.38	149	-5.67	149	-5.67	106	24.82	84	40.43
READ FOUR	62	72	-16.13	82	-32.26	56	9.68	54	12.90	55	11.29	80	-29.03	93	-50.00
CHI SQUARE			41.0981		237.9553		13.0889		11.8704		19.9001		32.9986		135.0869
STD. PREDICTION ERROR			25.9769		62.4580		14.7173		13.9786		18.0555		24.8918		48.6744
A-B-C-E															
READ NONE	622	610	1.93	589	5.31	644	-3.54	646	-3.86	642	-3.22	614	1.29	599	3.70
READ ONE	71	106	-49.30	150	-111.27	25	64.79	24	66.20	30	57.75	87	-22.54	121	-70.42
READ TWO	96	61	36.46	59	38.54	101	-5.21	98	-2.08	99	-3.13	95	1.04	84	12.50
READ THREE	114	126	-10.53	85	25.44	153	-34.21	153	-34.21	151	-32.46	99	13.16	79	30.70
READ FOUR	97	97	0.00	118	-21.65	77	20.62	78	19.59	78	19.59	105	-8.25	116	-19.59
CHI SQUARE			31.5086		115.8362		48.3072		49.1441		40.1433		6.3524		52.0290
STD. PREDICTION ERROR			23.4009		44.6789		30.1529		30.5647		27.6405		11.0454		30.8513
A-B-D-E															
READ NONE	634	627	1.10	604	4.73	644	-1.58	641	-1.10	643	-1.42	618	2.52	603	4.89
READ ONE	83	103	-24.10	149	-79.52	61	26.51	64	22.89	67	19.28	121	-45.78	156	-87.95
READ TWO	131	110	16.03	112	14.50	137	-4.58	139	-6.11	125	4.58	115	12.21	99	24.43
READ THREE	107	115	-7.48	67	37.38	121	-13.08	120	-12.15	131	-22.43	90	15.89	77	28.04
READ FOUR	45	45	0.00	68	-51.11	37	17.78	36	20.00	34	24.44	56	-24.44	66	-46.67
CHI SQUARE			8.8611		83.3660		9.5179		8.2947		11.5590		25.1454		91.7486
STD. PREDICTION ERROR			13.8130		39.3599		13.2665		12.0333		14.6287		21.7532		41.6053
A-C-D-E															
READ NONE	630	622	1.27	604	4.13	639	-1.43	637	-1.11	639	-1.43	612	2.86	597	5.24
READ ONE	85	105	-23.53	137	-61.18	60	29.41	62	27.06	62	27.06	121	-42.35	157	-84.71
READ TWO	120	110	8.33	123	-2.50	139	-15.83	140	-16.67	135	-12.50	117	2.50	101	15.83
READ THREE	124	117	5.65	68	45.16	124	0.00	124	0.00	127	-2.42	92	25.81	78	37.10
READ FOUR	41	46	-12.20	69	-68.29	38	7.32	37	9.76	37	9.76	58	-41.46	67	-63.41
CHI SQUARE			6.6457		77.3721		10.7094		10.0249		8.6899		31.1432		99.2775
STD. PREDICTION ERROR			11.2960		38.2335		14.6697		14.0996		13.1149		24.2570		43.4189
B-C-D-E															
READ NONE	626	610	2.56	594	5.11	636	-1.60	635	-1.44	636	-1.60	607	3.04	590	5.75
READ ONE	67	109	-62.69	147	-119.40	43	35.82	44	34.33	44	34.33	107	-59.70	147	-119.40
READ TWO	122	94	22.95	74	39.34	132	-8.20	133	-9.02	131	-7.38	114	6.56	98	19.67
READ THREE	130	125	3.85	114	12.31	141	-8.46	141	-8.46	141	-8.46	100	23.08	82	36.92
READ FOUR	55	62	-12.73	71	-29.09	48	12.73	48	12.73	48	12.73	71	-29.09	83	-50.91
CHI SQUARE			34.2468		122.6672		11.3981		10.8384		10.5409		35.5595		134.2916
STD. PREDICTION ERROR			23.9917		45.2548		13.7550		13.4239		13.2665		25.2230		47.6655

Test d'Ajustement de la distribution des Occasions de Voir sur la base des données de Marc

Combinaison de 5 supports	Distribution Observée	Kwerel		Sainsbury		Agré-Var.		Agré. Const.		Indé. Cond.		Hofmans		Metheringham	
A-B-C-D-E		EST	% E	EST	%E	EST	%E	EST	%E	EST.	%E	EST.	%E	EST	%E
READ NONE	616	590	4.22	560	9.09	634	-2.92	633	-2.76	635	-3.08	593	3.73	567	7.95
READ ONE	50	114	-128.00	166	-232.00	22	56.00	20	60.00	19	62.00	88	-76.00	136	-172.00
READ TWO	86	48	44.19	46	46.51	72	16.26	78	9.30	75	12.79	93	-8.14	89	-3.49
READ THREE	108	101	6.48	92	14.81	138	-27.78	135	-25.00	134	-24.07	92	14.81	71	34.26
READ FOUR	100	104	-4.00	68	32.00	104	-4.00	105	-5.00	107	-7.00	76	24.00	65	35.00
READ FIVE	40	43	-7.50	68	-70.00	30	25.00	29	27.50	29	27.50	59	-47.50	72	-80.00
CHI SQUARE		100.6468		325.0259		29.4784		29.2383		30.9873		47.4639		202.4483	
STD. PREDICTION ERROR		32.3780		58.1034		19.6636		18.8326		19.5320		23.1481		47.3005	

COMPARAISON DES PERFORMANCES DE 7 MODÈLES DE DUPLICATION SUR LA BASE DE LA DISTRIBUTION DES OCCASIONS DE VOIR *

(pour chaque combinaison de magazines et pour chaque méthode, la Table donne les valeurs du Test χ^2 et de l'erreur-type de prédiction. De faibles valeurs indiquent un bon ajustement des fréquences estimées aux fréquences observées)

Magazines	Tests	Kwerel	Sainsbury modifiée	Agrégation séquentielle	Goodhart (1)	Indépendance conditionnelle (1)	Hofmans étendue (1)	Metteringham
				A. COMBINAISON DE TROIS MAGAZINES				
A.B.C.	χ^2	2,7	92,4	21,7	21,9	19,1	2,9	20,5
	Erreur-type	8,6	50,8	24,6	24,6	23,0	8,9	23,9
A.B.D.	χ^2	0,1	7,2	1,6	1,4	4,8	7,8	57,5
	Erreur-type	1,6	15,7	7,4	6,7	12,7	16,3	35,8
A.B.E.	χ^2	0,0	34,5	11,0	11,5	11,0	3,1	21,8
	Erreur-type	0,7	32,9	18,6	19,0	18,6	11,2	26,2
A.C.D.	χ^2	1,5	30,8	0,2	0,6	0,0	24,1	91,1
	Erreur-type	7,4	32,9	2,9	4,7	0,7	29,1	46,3
A.C.E.	χ^2	1,4	5,1	28,1	28,8	22,9	0,3	9,4
	Erreur-type	6,7	12,7	29,7	30,2	26,8	3,0	17,2
A.D.E.	χ^2	6,0	74,7	0,1	0,1	0,2	9,6	24,3
	Erreur-type	14,1	49,9	1,6	1,2	3,0	17,9	28,4
B.C.D.	χ^2	10,7	51,9	1,1	0,8	1,7	18,7	74,9
	Erreur-type	18,6	40,9	6,0	5,3	7,4	24,6	49,2
B.C.E.	χ^2	2,7	18,2	19,6	21,1	13,9	4,5	27,6
	Erreur-type	8,9	23,0	23,9	24,7	20,1	11,2	28,4
B.D.E.	χ^2	4,0	11,6	3,2	2,8	3,6	7,0	21,7
	Erreur-type	11,9	20,1	10,5	9,6	11,2	15,7	27,5
C.D.E.	χ^2	0,8	18,0	0,1	0,5	0,0	21,5	55,5
	Erreur-type	5,1	24,6	2,2	4,1	0,7	26,8	43,2
				B. COMBINAISON DE QUATRE MAGAZINES				
A.B.C.D.	χ^2	41,1	238,0	13,1	11,9	19,9	53,0	135,1
	Erreur-type	26,0	62,5	14,7	14,0	18,1	24,9	48,7
A.B.C.E.	χ^2	31,5	115,8	48,3	49,1	40,1	6,4	52,0
	Erreur-type	23,4	44,7	30,2	30,6	27,6	11,0	30,9
A.B.D.E.	χ^2	8,9	8,3	9,5	9,3	11,6	25,1	91,7
	Erreur-type	13,8	39,4	13,3	12,0	14,6	21,8	41,6
A.C.D.E.	χ^2	6,6	77,4	10,7	10,0	8,7	31,1	99,3
	Erreur-type	11,3	38,2	14,7	14,1	13,1	24,3	43,3
B.C.D.E.	χ^2	34,2	122,4	11,4	10,8	10,3	36,6	134,3
	Erreur-type	24,0	45,4	13,8	13,4	13,3	25,2	47,7
				C. COMBINAISON DE CINQ MAGAZINES				
A.B.C.D.E.	χ^2	100,6	325,0	29,3	29,2	31,0	47,5	202,4
	Erreur-type	32,4	58,1	19,7	18,9	19,5	23,1	47,3

* Données de MARC.

(1) Modèles développés par l'auteur.

Commentaires

Les résultats des tables précédentes montrent que le modèle de Metheringham obtient de meilleurs résultats pour les données de Politz que pour les données de Marc. Le modèle de Metheringham surestime systématiquement la couverture nette mais cette surestimation est stable. Elle ne s'accroît pas en fonction du nombre de supports inclus dans la combinaison. Le pourcentage d'erreurs n'est que de 1,5% en moyenne pour la combinaison de 11 supports, ce qui constitue le meilleur résultat. Pour les données de Politz, le modèle de Metheringham est supérieur au modèle d'Hofmans et au modèle d'Agostini, et il est très supérieur à tous les autres modèles stochastiques. Pour les données de Marc, le modèle de Metheringham est légèrement inférieur au modèle d'Hofmans et à la plupart des modèles stochastiques à l'exception du modèle de Sainsbury modifié et du modèle binomial.

En ce qui concerne l'estimation de la distribution des occasions de voir, le modèle de Metheringham est handicapé par le fait que la distribution Bêta Binomiale ne possède jamais de pics intermédiaires. Il en résulte un lissage néfaste de la distribution des occasions de voir. Le nombre de concordance au niveau des pics secondaires est nul, les valeurs du test d'ajustement du Chi Deux sont supérieures à celles observées pour les autres modèles, à l'exception du modèle de Sainsbury modifié.

30. Le Modèle de Benson

Ce modèle est certainement le plus ambitieux de tous ceux passés en revue puisqu'il combine l'hétérogénéité des individus avec celle des supports et avec l'interdépendance des audiences.

Benson (26) assume que les probabilités individuelles d'exposition à chaque support suivent une distribution normale, et que la distribution des probabilités d'exposition à une paire de supports suit une loi normale à deux dimensions. L'interdépendance des audiences de deux supports peut alors être mesurée par leur coefficient de corrélation.

Dans le cas de deux supports, le coefficient de corrélation peut être évalué aisément à partir d'un tableau croisé à 2 dimensions du nombre de personnes exposées. Par contre, dès que l'on considère plus de 2 supports on se heurte à des difficultés de calcul non négligeables.

Par ailleurs, l'hypothèse de normalité qui implique que le plus grand nombre d'individus se situent aux alentours de la probabilité moyenne d'exposition, ne correspond pas aux observations empiriques. Celles-ci font apparaître, au contraire, une concentration des individus aux deux extrémités de l'échelle des probabilités.

Pour ces deux raisons le modèle de Benson est davantage une hypothèse d'école qu'un modèle réellement opérationnel.

Nous avons terminé notre étude des modèles de duplication. Nous sommes maintenant en mesure d'énumérer les divers critères d'évaluation qu'il convient de considérer pour établir un classement des modèles de duplication. La liste de ces critères,qui dérivent de considérations théoriques, empiriques et pratiques,est donnée dans la table suivante. Nous appliquons ensuite ces 15 critères aux 14 modèles de duplication quin ont été étudiés tout au long de ce chapitre. Cette seconde table résume, de façon synthétique, les résultats de notre analyse des modèles de duplication.

15 Critères d'Evaluation pour les Modèles de Duplication d'Audience

Les Critères théoriques

1 - L'exposition aux supports est-elle considérée comme un phénomène déterminé ou comme un phénomène aléatoire ?

2 - Les individus composant la population sont-il considérés comme homogènes dans leur comportement d'exposition ?

3 - Les supports sont-il considérés comme homogènes ?

4 - Le phénomène de duplication se produit-il au pur hasard ?

5 - Les paramètres du modèle deviennent-ils indéfinis ou infinis dans certains cas ?

6 - L'erreur d'estimation varie-t-elle en fonction de l'ordre d'inclusion des supports ?

7 - Le modèle produit-il des résultats logiquement inconsistants ? (c'est-à-dire diminution de la couverture nette après adjonction d'un support, audiences négatives pour certaines classes de fréquence.

Les critères empiriques

8 - L'estimation de la couverture nette est-elle systématiquement biaisée ?

9 - L'erreur d'estimation varie-t-elle en fonction du nombre de supports inclus ?

10 - La valeur absolue des erreurs est-elle inférieure à l'erreur d'échantillonnage dans la mesure des audiences ?

11 - Le modèle permet-il l'estimation de la distribution des occasions de voir ?

12 - Le modèle reproduit-il fidèlement la distribution des occasions de voir ?

Les Critères pratiques

13 - Le modèle nécessite-t-il des calculs volumineux ?

14 - Le volume des calculs est-il fonction du nombre de supports inclus ?

15 - Le modèle nécessite-t-il des données d'audience non habituellement disponibles ?

Evaluation Multicritères de 14 Modèles de Duplication

Modèles	Critères (1) Théoriques							Empiriques					Pratiques		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Modèles Ad-Hoc															
Agostini	D	NA	Y	C	N	N	N	O	Y	N	N	NA	N	N	N
Hofmans	D	NA	N	N	N	N	N	O	N	N	Y	NA	N	N	N
Starch	D	NA	N	N	N	Y	N	U	Y	N	N	NA	Y	Y	N
Starch modifié	D	NA	N	N	N	N	N	U	Y	N	N	NA	Y	Y	N
Little & Lodish	D+S	Y	Y	C	N	N	?	O	?	Y	N	N	N	N	N
Modèles stochastiques															
Modèle Binomial	S	Y	Y	Y	N	N	N	O	Y	Y	N	N	N	N	N
Sainsbury	S	Y	N	Y	N	N	N	O	Y	Y	N	N	N	N	N
Kwerel	S	Y	Y	N	?	N	N	U	Y	Y	N	Y	Y	Y	N
Sainsbury Modified	S	Y	N	N	N	N	Y	O	Y	Y	N	N	Y	Y	N
S.Aggreg. w. Cst. Dup.	S	Y	N	C	N	Y	N	U	Y	Y	N	N	N	N	N
S. Aggreg. w. Var. Dup.**	S	Y	N	N	N	Y	N	U	Y	Y	N	N	N	N	N
Indépendance conditionnelle	S	Y	N	N	N	N	N	U	N	Y	N	N	Y	Y	N
Hofmans Etendu	S	Y	N	N	N	N	N	O	N	Y	Y	N	N	N	N
Metheringham	S	N	Y	N	Y	Y	N	O	Y	Y	Y	N	N	N	N

Signification des Symboles

NA = Le critère ne s'applique pas

Y = Oui

N = Non

? = Réponse inconnue.

Pour le critère n° 1 : D = Déterministe.
S = Stochastique

Pour le critère n° 4 : C = La duplication est constante.

Pour le critère n° 8 = O = Surestimation.
U = Sous-estimation.

(1) Voir liste des critères page précédente.

** Modèles développés par l'auteur.

3. Evaluation et Classement des Modèles de Duplication

Il ressort de notre étude une classement en trois groupes, par ordre de mérite croissant, des modèles de duplication d'audience :

- le premier groupe rassemble trois modèles qui sont systématiquement dominés tant sur le plan de la couverture nette que sur celui de la distribution des occasions de voir : le modèle Binomial, le modèle de Sainsbury, et, dans une moindre mesure, le modèle de Sainsbury Modifié. Tous trois surestiment la couverture nette et se montrent incapables de reproduire correctement la distribution des occasions de voir, quel que soit le nombre de supports;

- le deuxième groupe comprend les modèles qui obtiennent des résultats satisfaisants pour un petit nombre de supports, mais sous-estiment gravement la couverture nette lorsque le nombre de supports s'accroît. Les deux versions du modèle de Starch et du modèle d'agrégation séquentielle font partie de ce groupe, ainsi que le modèle de l'indépendance conditionnelle et le modèle de Kwerel.

Tous ces modèles démontrent une bonne capacité à reproduire la distribution des occasions de voir ainsi que la couverture nette lorsque 5 supports où moins sont inclus. Par contre, leurs prestations décroissent et le volume des calculs augmente drastiquement dès que le nombre de supports dépasse 5.

- Le troisième groupe comprend trois modèles qui performent bien pour les combinaisons incluant un grand nombre de supports et moins bien lorsque le nombre de supports diminue : le modèle d'Agostini, le modèle de Metheringham et le modèle d'Hofmans Modifié. Ce dernier est le meilleur des trois car ses performances ne sont guères influencées par le nombre des supports inclus, et car il est celui des trois qui reproduit le mieux la distribution des occasions de voir.

Les deux tables suivantes présentent une comparaison d'ensemble des résultats pour les 3 groupes de modèles, tant du point de vue de l'estimation de la couverture nette que du point de vue de l'estimation de la distribution des occasions de voir.

Comparaison des Performances de 11 Modèles de Duplication

Estimation de la Couverture Nette

(Erreur moyenne en pourcentage de la Couverture Observée)

	Données de Marc			Données de Politz	
Nombre de Magazines	3	4	5	5	11
Nombre de Tests	10	5	1	3	6
Modèles					
Groupe 1					
Binomial, Sainsbury	+ 52,9 %	+ 73,7 %	+ 90,6 %	+ 18,8 %	+ 24,6 %
Sainsbury Modifié	+ 3,7 %	+ 8,8 %	+ 14,6 %	+ 6,2 %	+ 12,7 %
Groupe 2					
Starch	+ 1,8 %	- 3,7 %	- 6,3 %	- 6,0 %	- 20,9 %
Starch modifié *	+ 1,0 %	- 2,1 %	- 1,3 %	- 4,4 %	- 16,6 %
Agregation Séquent. avec duplication constante	- 1,8 %	- 2,9 %	- 4,4 %	- 5,7 %	- 29,1 %
Agrégation Séquent. avec duplication variable *	- 1,7 %	- 3,3 %	- 4,7 %	- 5,1 %	- 19,4 %
Indépendance Conditionnelle *	- 3,4 %	+ 3,2 %	- 4,9 %	- 5,0 %	N.A.
Kwerel	+ 1,0 %	+ 3,3 %	+ 6,8 %	- 4,2 %	- 13,5 %
Groupe 3					
Agostini est.	+ 14,5 %	+ 20,4 %	+25,3 %	+ 5,2 %	+ 3,9 %
Hofmans	+ 2,1 %	+ 4,2 %	+ 6,0 %	+ 1,5 %	+ 2,5 %
Metheringham	+ 4,1 %	+ 8,5 %	+12,8 %	+ 1,1 %	+ 1,5 %

* Méthodes développées par l'auteur.

Comparaison des Performances de 7 Modèles de Duplication

Estimation de la Distribution des Occasions de Voir

(Chi Deux Moyen)

	Données de Marc		
Groupe 1	3 supports	4 supports	5 supports
Binomial			
Sainsbury Modifié	34,44	112,38	325,0
Groupe 2			
Agrégation Séquentielle avec duplication constante	8,95	18,22	29,2
Agrégation Séquentielle avec duplication variable	8,67	18,60	29,5
Indépendance Conditionnelle	7,01	18,16	31,0
Kwerel	2,99	24,27	100,65
Groupe 3			
Hofmans Etendu	9,93	26,44	47,46
Metheringham	35,43	102,48	202,45

4 - Conclusions.

Dans ce chapitre, nous avons étudié six modèles de duplication du type Ad Hoc et dix modèles stochastiques de duplication. Dans la première catégorie de modèles, nous avons présenté la formule d'Agostini et examiné son principal défaut, la constance de k. Après avoir discuté les diverses hypothèses implicites sous-jacentes à la formule d'Agostini, nous avons étudié la formule d'Hofmans qui tient compte des variations de duplication d'une paire de supports à une autre. Nous avons montré que la formule d'Hofmans contient une hypothèse implicite concernant les recouvrements d'ordre supérieur à deux. Nous avons ensuite étudié le modèle de Starch et nous avons démontré que la couverture nette estimée dépend de l'ordre d'inclusion des supports. Nous avons proposé une amélioration de la formule de Starch en introduisant un ordre optimal d'inclusion. Après avoir présenté la formule de Little et Lodish, nous avons fait apparaître la très mauvaise qualité de ses estimations.

En ce qui concerne les modèles stochastiques de duplication, nous avons présenté une classification générale de ces modèles en fonction des hypothèses faites sur l'homogénéité des individus et des supports et sur l'interdépendance des audiences. Nous avons montré que les deux modèles basés sur l'hypothèse de l'indépendance des audiences conduisent à une surestimation systématique et inacceptable de la couverture nette. Nous avons établi que les deux meilleurs modèles sont ceux d'Hofmans et de Metheringham qui tiennent compte explicitement de la corrélation des audiences. Tous les autres modèles stochastiques conduisent à des estimations dont la qualité se dégrade rapidement dès que le nombre de supports est supérieur à cinq.

Finalement, nous avons construit une grille d'évaluation comportant 15 critères qui résument les principales caractéristiques que devrait posséder un modèle de duplication satisfaisant. Nous avons utilisé cette grille pour faire apparaître d'une part les modèles systématiquement dominés et, d'autre part, les meilleurs modèles actuellement disponibles.

Notre principale contribution réside dans le rapprochement de 15 modèles de duplication et dans la réalisation de la première comparaison exhaustive de leurs résultats.

Notre seconde contribution réside dans l'amélioration que nous avons apportée au modèle d'Hofmans en l'étendant à l'estimation de la distribution des occasions de voir.

Notre troisième contribution réside dans la mise au point d'une grille d'évaluation multicritères permettant d'évaluer les performances d'ensemble de chaque modèle.

o o
o

NOTES

(1) - L'enquête du CESP de 1957, qui a été tabulée par toutes les combinaisons de 15 magazines parmi 30, remplit 3 énormes volumes. Une tabulation similaire des 105 magazines actuellement mesurés par T.G.I. remplirait plusieurs bibliothèques.

(2) - Par exemple la tabulation complète de tous les 44 magazines mesurés dans l'enquête d'audience Allemande de 1961 comprendrait un total de 17,6 milliards de combinaisons différentes

(3) - AGOSTINI J.M. , "How to estimate undiplicated audiences", Journal of Advertising Research, Vol.1, March 1961, pp. 11-14

(4) - Centre d'Etude des Supports de Publicité, 8, rue Danielle-Casanova, Paris

(5) - KUHN W. ,"Net audiences of German magazines" Journal of Advertising Research, Vol. 3, March 1963, pp. 30-33

(6) - CAFFYN J.M., SAGOVSKY M., "Net audiences of British newspapers : a comparison of the Agostini and Sainsbury methods", Journal of Advertising Recherch, Vol.3, March 1963, pp. 21-24

(7) - MARC M., "Net Audiences of French business papers : Agostini formula applied to special markets", Journal of Advertising Research, Vol.3, March 1963, pp. 26-29

(8) - BOWERS J. , "Net audiences of U.S. and Canadian magazines : seven tests of Agostini formula", Journal of Advertising Research, Vol.3, March 1963, pp. 13-20

(9) - KAATZ R., "Improving Agostini's formula for net audiences", Journal of Advertising Research, Vol.3, September 1963, pp.43-44

(10) - MARC M. , "Problèmes posés par les duplications d'audience entre titres : distribution des répétitions", Etude IREP N°6, Novembre 1962

(11) - HOFMANS P., "Measuring the cumulative net coverage of any combination of media", Journal of Marketing Research, Vol.3, August 1966, pp. 269-278

(12) - CLAYCAMP H.J., Mc Clelland C.W., "Estimating reach and the magic of k", Journal of Advertising Research, Vol.8, N°2, pp.44-51

(13) - Pour un complément d'information sur ce sujet, consulter : Fréchet M., "Les probabilités associées à un système d'évènements compatibles et dépendants", Actualités Scientifiques et Industrielles, N° 859 et 942, Paris, 1940 et 1943

(14) - Pour la démonstration du théorème de Waring on pourra consulter : Feller W., An Introduction to Probability Theory and its Application Vol.1, p.106, J. Wiley & Sons, New York

(15) - STARCH & STAFF, Inc., Starch Magazine Report, 1967, New York

(16) - YOUNG L.F., "Estimating Radio Reach", Journal of Advertising Research, Vol.12, October 1972, pp.37-41

(17) - LITTLE J.D.C., LODISH L.M. "A media planning calculus" Operation Research, January 1969, pp. 3-35

(18) - POLITZ A. Media Studies, Main Report - 1964 Magazine Study - Audiences of 11 Magazines, Etude financée par Look, Mc Call's, Reader Digest, Redbook, The Saturday Evening Post, February 1965, New York

(19) - Même référence que 17 : " ... jusqu'à présent nous n'avons pu trouver une distribution qui soit facile à calculer et qui reproduise suffisamment bien les données."

(20) - ELLIS D.M. "Building up a sequence of optimum media schedules", Operational Research quaterly, Vol.17 , N°4, pp.413-423

(21) - KWEREL S.M., "Estimating Unduplicated Audience and Exposure Distribution" Journal of Advertising Research, Vol. 9, June 1969 pp. 46-53

(22) - KWEREL S.M., "Information retrieval for media planning", Management Science, Vol. 15, N°4, Décembre 1968, pp.B137-B160

(23) - KWEREL S.M., "Estimating the undiplicated audience of a combination of media vehicles : integrated theory and estimation methods", Ph.D. Dissertation, Columbia University, 1964

(24) - GOODHART G.J. & A.S.C. EHRENBERG, "Duplication of Television Viewing Between and Within Channels, Journal of Marketing Research May 1969, pp. 169-178.

(25) - METHERINGHAM R.A. " Measuring the net cumulative coverage of a print Campaign", Journal of Advertising Research, December 1964 pp. 23-28

(26) - BENSON P.H. "Bivariate Normal Distribution to Calculate Media Exposure", Journal of Advertising Research, Vol. 9 N°3 1969, pp. 41-47.

CHAPITRE 7

LES MODELES D'EXPOSITION COMPLETS

Le problème auquel les modèles d'exposition complets s'efforcent d'apporter une réponse peut s'énoncer comme suit : lorsqu'un plan publicitaire comporte de multiples insertions dans de multiples supports, combien d'individus appartenant à la population cible seront-ils exposés et combien de fois ?

Nous étudierons d'abord la relation liant l'accumulation d'audience à la distribution des occasions de voir, puis nous analyserons les divers modèles stochastique et de simulation permettant d'estimer la couverture nette et la distribution des occasions de voir engendrées par des plans publicitaires à insertions multiples dans plusieurs supports.

1 - La relation entre l'accumulation d'Audience et la Distribution des Occasions de voir : généralisation à plusieurs supports.

La relation mise à jour entre l'accumulation d'audience et la fréquence d'exposition, dans le cas d'un seul support, peut être généralisée au cas de plusieurs supports. Afin de faciliter la compréhension du mécanisme de cette relation, il est utile d'examiner tout d'abord un exemple numérique ne comportant que deux supports.

La classification croisée des résultats de l'enquête d'audience menée en 1973 par Simmons permet d'obtenir les matrices d'exposition, à une et à deux parutions, des magazines Time et Newsweek. Ces deux matrices d'exposition ont été présentées au chapitre 4, pages 121 et 122 . Considérons d'abord la relation qui lie la matrice d'exposition à deux parutions à la matrice d'exposition à une

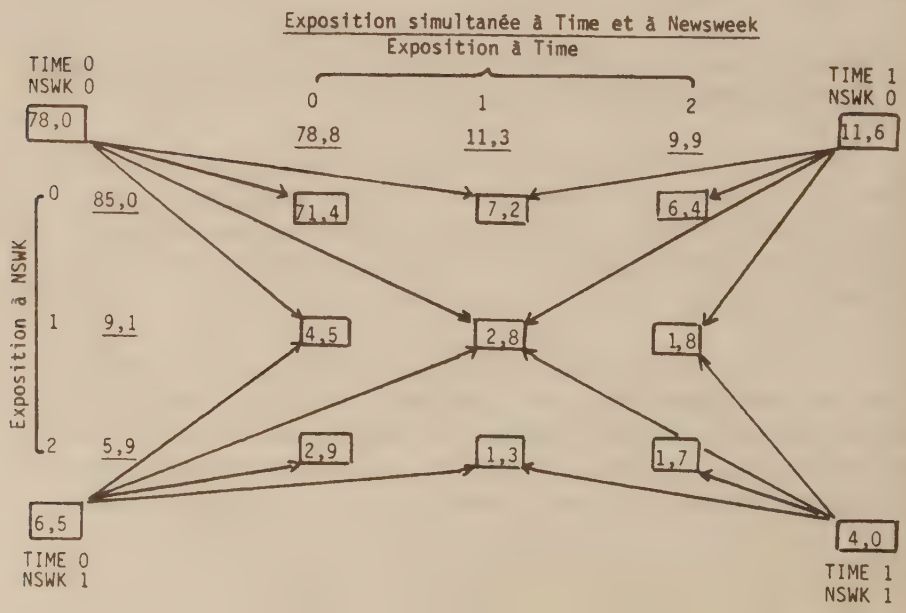
parution.

11 - Une relation récursive pour les matrices d'exposition.

Lorsque deux parutions sont considérées, les individus qui ne sont exposés ni à la première parution de Time, ni à la première parution de Newsweek, appartiennent à l'une des quatre catégories suivantes :

- ceux qui ne sont exposés à aucune parution de Time ni de Newsweek ;
- ceux qui sont exposés à la seconde parution de Time et à la seconde de Newsweek ;
- ceux qui sont exposés à la seconde parution de Time et ne le sont pas à la seconde de Newsweek ;
- ceux qui sont exposés à la seconde parution de Newsweek et ne le sont pas à la seconde de Time.

La même analyse peut être pratiquée pour les individus exposés à la première parution de Time et à la première de Newsweek et ainsi de suite. Le tableau suivant décrit comment l'on déduit, à partir des quatre cases de la matrice d'exposition à une parution, les neuf cases de la matrice d'exposition à deux parutions.



D'après ce tableau, on peut écrire des équations qui donneront les valeurs des quatre cases de la matrice à une parution en fonction des valeurs de la matrice à deux parutions.

Adoptons la notation suivante :

$M = \{1, 2, \ldots. i, j, \ldots. m\}$ = Ensemble des m supports

$S = \{n1, n2, \ldots. ni, nj, \ldots. nm\}$ = Plan d'insertions avec n1 insertions dans le support 1, n2 dans le support 2 , etc...

$K_i = \{k1, k2, \ldots ki, kj, \ldots km\}$ = Vecteur d'exposition de l'individu i : l'individu i est exposé à k1 parutions du support 1, k2 parutions au support 2, ... et ainsi de suite.
Notons que ki vérifie la relation : $0 \leq ki \leq ni$;entiers.

$E^{ni,\,nj}_{ki,\,kj}$ = le nombre d'individus exposés à ki parutions du support i, kj parutions du support j, quand il y a ni parutions dans le support i et nj parutions dans le support j. Nous utiliserons cette notation pour les valeurs absolues ou les pourcentages de la population totale.

$R_{ni,\,nj}$ = le nombre (ou pourcentage) d'individus exposés à au moins une parution du support i et à au moins une parution du support j, quand il y a ni insertions dans le support i et nj insertions dans le support j.

Quand nous traiterons le cas de plus de deux supports, des indices supplémentaires seront ajoutés selon leur nécessité. Avec ces notations, et dans l'hypothèse d'une audience constante d'une parution à une autre pour les deux supports, les équations qui convertissent la matrice d'exposition à deux parutions en une matrice d'exposition à une parution sont les suivantes :

$$(1)\quad \begin{cases} E^{11}_{00} = E^{22}_{00} + 1/2E^{22}_{01} + 1/2E^{22}_{10} + 1/4E^{22}_{11} \\ E^{11}_{01} = E^{22}_{02} + 1/2E^{22}_{01} + 1/2E^{22}_{12} + 1/4E^{22}_{11} \\ E^{11}_{10} = E^{22}_{20} + 1/2E^{22}_{10} + 1/2E^{22}_{21} + 1/4E^{22}_{11} \\ E^{11}_{11} = E^{22}_{22} + 1/2E^{22}_{12} + 1/2E^{22}_{21} + 1/4E^{22}_{11} \end{cases}$$

Un système d'équations similaire peut être dérivé pour convertir une matrice d'exposition à 3 parutions en une matrice à deux parutions. Un tel système comprendrait 9 équations. Toutefois, il existe un moyen plus simple pour générer les matrices d'exposition de rang inférieur. Il consiste en la réduction séquentielle du nombre de parutions, un support à la fois,en utilisant l'équation de récurrence (33) du chapître 5, page 194, appliqué séparément à chaque colonne (ou chaque ligne) de la matrice.

La formule générale de récurrence qui en résulte est :

$$(2) \quad E^{ni,\,nj-1}_{ki,\,kj} = \left(\frac{ni - kj}{nj}\right) E^{ni,\,nj}_{ki,\,kj} + \left(\frac{kj + 1}{nj}\right) E^{ni,\,nj}_{ki,\,kj+1}$$

pour $kj = 0, 1, \ldots, nj - 1$; pour $ki = 0, 1, \ldots, ni$

Par exemple, une matrice de dimension 2 x 2 peut être réduite à une matrice de dimension 2 x 1, en appliquant la formule suivante aux 3 lignes de la matrice initiale :

$$E^{2,1}_{i,j} = 1/2E^{2,2}_{i,j} + E^{2,2}_{i,\,j+1} \quad \text{pour } j = 0,1 \text{ et } i = 0, 1, 2$$

Alors que le format de la matrice d'exposition possède l'avantage de la concision, il masque cependant le fait que l'on peut atteindre n'importe quelle cas d'exposition par divers itinéraires. Par conséquent, il sera parfois plus pratique d'utiliser le format de la distribution jointe d'exposition où toutes les séquences d'exposition sont conservées comme telles. La table suivante éclaire le lien entre les deux formats.

12 - <u>Conversion des matrices d'exposition en matrices de couverture nette</u>.

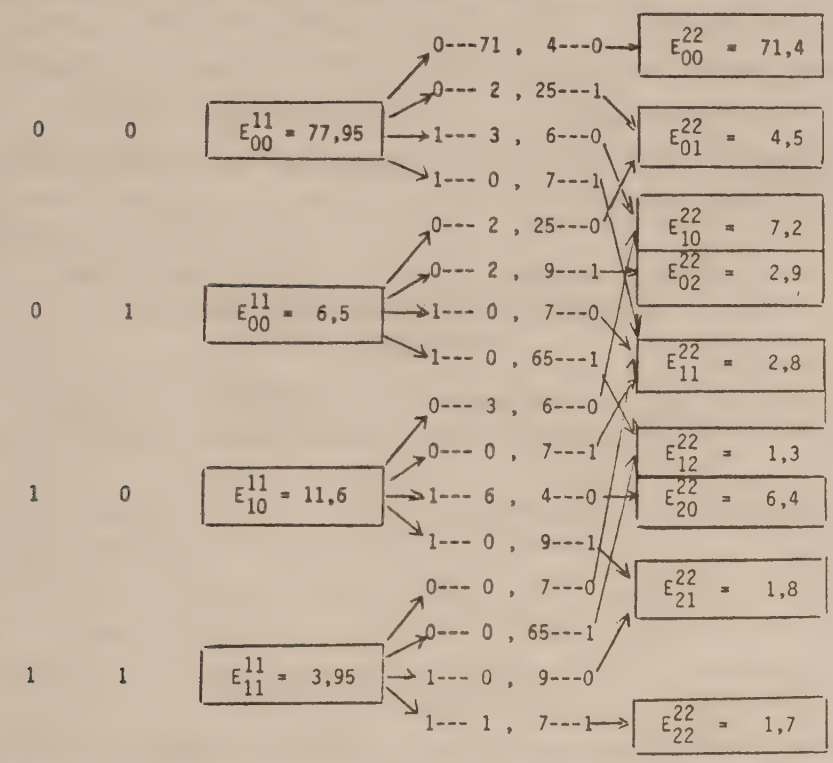
Nous avons déjà traité d'une manière détaillée, dans le chapître 5, page et suivantesde la manière dont le théorème hypergéométrique peut être utilisé pour convertir la distribution des occasions de voir en couverture nette cumulée pour un support. L'extension au cas de plusieurs supports est relativement directe. Cependant, deux problèmes sont à distinguer :

- Premièrement, on peut avoir besoin d'obtenir la couverture nette cumulée d'un support, conditionnelle au nombre d'expositions à un autre support ;

- Deuxièmement, il peut être intéressant de connaître la couverture

Distribution jointe d'exposition à Time et Newsweek 1973

(1 signifie exposition, 0 signifie non-exposition)



nette cumulée de toutes les combinaisons possibles de parutions qui peuvent être formées à partir d'une matrice d'exposition donnée.

Commençons par le premier problème. La solution consiste à appliquer la formule (1) de la page 415 à chaque colonne, séparément ou à chaque ligne de la matrice d'exposition. Dans le cas de trois supports ou plus, nous avons à faire à une matrice à m dimensions et la procédure consiste en la conversion de chaque vecteur dans la ième dimension sélectionnée. L'exemple de calcul suivant illustre la procédure, en convertissant une matrice d'exposition Time-Newsweek à deux parutions, en deux matrices de couverture nette cumulées conditionnelles.

Conversion des colonnes de la matrice d'Exposition en colonnes de la Matrice de couverture nette cumulative

Expositions	Coefficients hypergéométr. R_1	R_2	Colonne 1 Exposit.	Colonne 1 Couvert.	Colonne 2 Exposit.	Colonne 2 Couvert.	Colonne 3 Exposit.	Colonne 3 Couvert.
0	0	0	95 355	-	9 591	-	8 525	-
1	1/2	1	5 978	6 798	3 728	3 636	2 421	3 550
2	1	1	3 809	9 787	1 772	5 500	2 340	4 761

Par exemple, la couverture nette cumulée de la première colonne est obtenue comme suit :

$$R_1 = 1/2 * 5\,978 + 3\,809 = 6\,798$$

$$R_2 = 5\,978 + 3\,809 = 9\,787$$

Le résultat de la conversion est une matrice qui indique la couverture nette cumulée de Newsweek en fonction de la fréquence d'exposition à Time.

Couverture nette de Newsweek en fonction de la fréquence d'exposition à Time

Couverture nette cumulée de Newsweek	Exposition à Time 0	1	2	Couverture nette totale de Newsweek
1 parution	6 798	3 636	3 550	13 984
2 parutions	9 787	5 500	4 761	20 048

Naturellement, nous aurions pu commencer en convertissant d'abord les colonnes. Nous aurions alors obtenu la couverture nette cumulée de Time en fonction de la fréquence d'exposition à Newsweek.

La couverture nette de Time en fonction de la fréquence d'exposition à Newsweek

Couverture nette cumulée de Time

		Une parution	Deux parutions
Expositions à Newsweek	0	13 321	18 116
	1	4 285	6 149
	2	3 226	4 112
Couverture nette de Time		20 832	28 377

Considérons maintenant le second problème. Nous voulons convertir la matrice d'exposition à deux parutions Time-Newsweek en matrice de couverture nette cumulée à deux parutions Time-Newsweek. La manière la plus rapide est d'utiliser la formule :

$$(3) \qquad R_{ni,nj} = POP - E^{ni,\,nj}_{0,0}$$

On peut faire cela en deux étapes. D'abord, la matrice d'exposition à deux parutions est réduite à la matrice d'exposition ni, nj désirée, en utilisant la formule de récurrence (2), puis on applique ensuite la formule (3). La mise en oeuvre de cette procédure pour nos données fournit la matrice de couverture nette cumulée suivante :

Matrice de couverture nette cumulée à deux parutions, Time-Newsweek

		TIME		
		Aucune PARUTION	Une PARUTION	Deux PARUTIONS
NEWSWEEK	Aucune PARUTION	0	20 832	28 377
	Une PARUTION	13 984	29 448	35 176
	Deux PARUTIONS	22 048	33 370	38 165

Cette table donne la couverture nette des neuf plans d'insertions qui peuvent être construits quand on considère deux parutions au plus pour chaque support. Si la couverture nette est le seul critère considéré dans l'élaboration d'un plan publicitaire, cette matrice fournit toutes les informations nécessaires. Par exemple, si le choix est limité à une insertion, il est préférable de l'avoir dans Time. Si deux insertions sont considérées, il vaut mieux en avoir une dans chaque revue. Si trois insertions sont considérées, il vaut mieux en avoir deux dans Time et une dans Newsweek, et ainsi de suite.

Toutefois, les données de l'enquête ne fournissent habituellement que la couverture nette cumulée de deux parutions pour chaque support et la couverture nette de toutes les paires de supports. Donc, l'objectif de tous les modèles d'exposition complets est d'estimer, à partir de ces données, toutes les matrices d'exposition et de couverture nette de rang supérieur à 2. Avant d'affronter ce problème, il convient de remarquer que la conversion inverse des matrices de couverture nette cumulée en matrices d'exposition est aussi possible.

13 - Conversion des matrices de couverture nette en matrices d'exposition.

Cette conversion s'obtient par une procédure similaire à celle de la conversion directe. Cette fois, nous devons utiliser la formule hypergéométrique

inverse (2) indiquée au chapître 5, page 140 . En convertissant la ligne 0 parution et la colonne 0 parution de la matrice de couverture nette, nous obtiendrons la ligne total et la colonne total de la matrice d'exposition. Les calculs sont présentés dans la table suivante.

Calcul de la ligne et de la colonne total de la matrice d'exposition

Parutions	Coefficients hypergéométriques inverses 1	2	Time Couverture nette	Fréquence Time 1	2	Newsweek Couverture nette	Fréquence Newsweek 1	2
1	-2	2	20 832	-41 664	+ 41 664	13 984	-27 068	+27 968
2	+2	-1	28 377	+56 754	- 28 377	20 048	+40 096	-20 048
				=15 090	= 13 287		=12 128	= 7 920

Ensuite, nous pouvons convertir la première ligne et la première colonne de la matrice de couverture nette. Cela nous donne les matrices d'exposition 2 par 1 et 1 par 2, comme nous l'indiquons dans la table suivante.

Calcul des matrices d'exposition 2 par 1 et 1 par 2

Parutions	Coefficients hypergéométriques inverses 1	2	Ligne 1 Couverture nette	Ligne 1 Fréquence d'exposition 1	2	Col. 1 Couverture nette	Col. 1 Fréquence d'exposition 1	2
1	-2	2	29 448	-58 896	+58 896	29 448	-58 896	+58 896
2	+2	-1	35 176	+70 352	-35 176	33 370	+66 740	-33 370
				=11 456	=23 720		= 7 844	=25 526
					-13 984			-20 832
					= 9 736			= 4 694

Deux parutions de Time
Une parution de Newsweek

	0 Time	1 Time	2 Time	Total
0 NWSK	98 344	11 456	9 736	119 356
1 NWSK	6 799	3 634	3 551	13 394
Total	105 143	15 090	13 287	133 520

Deux parutions de Newsweek
Une parution de Time

	0 Time	1 Time	Total
0 NWSK	100 150	13 322	113 472
1 NWSK	77 844	4 284	12 128
2 NWSK	4 694	3 226	7 920
Total	112 688	20 832	133 520

Finalement, en utilisant la deuxième ligne et colonne de la matrice de couverture nette, nous obtenons la matrice d'exposition à deux parutions.

Calcul de la matrice d'exposition à deux parutions

Parution	Coefficients Hypergéométriques inverses 1	2	Ligne 2 Couverture nette	Ligne 2 Fréquence d'exposition 1	2	Col. 2 Couverture nette	Col. 2 Fréquence d'exposition 1	2
1	- 2	2	33 370	-66 740	+66 740	35 176	-70 352	+70 352
2	+ 2	-1	38 165	+76 330	-38 165	38 165	+76 330	-38 165
					= 9 590	=28 575	= 5 978	=32 187
						-20 048		=28 377
						= 8 527		= 3 810

Matrice d'exposition à deux parutions Time-NWSK

	0 Time	1 Time	2 Time	Total
0 Newsweek	95 355	9 590	8 527	113 471
1 Newsweek	5 978	3 732	2 418	12 128
2 Newsweek	3 810	1 768	2 342	7 920
Total	105 142	15 090	13 287	135 519

Nous avons ainsi montré que les relations qui existent entre l'accumulation d'audience et la fréquence d'exposition dans le cas d'un seul support, peuvent être étendues au cas de plusieurs supports.

La conversion, directe ou inverse, requiert des données, soit sur la couverture nette cumulée, soit sur la distribution des occasions de voir. Ces données peuvent être obtenues par des classifications croisées des audiences des supports. Toutefois, ces tabulations exigent des manipulations fort coûteuses des banques de données et elles sont nécessairement limitées à deux parutions puisque ce sont là les seules données recueillies. Par conséquent, le spécialiste de la planification des supports doit recourir à des modèles d'exposition complets pour estimer la couverture nette et la distribution des occasions de voir de tout plan d'insertion prévoyant plus de deux parutions dans un même support. La suite de ce chapître analyse les divers modèles d'exposition qui ont été développés dans ce but. Nous discuterons leurs hypothèses puis nous évaluerons leurs performances sur la base de deux larges banques de données.

La première série de données, qui sera utilisée tout au long de ce chapître, comprend 57 plans publicitaires qui couvrent un large éventail de combinaisons de magazines. Les données d'audience proviennent de l'enquête Simmons de l'année 1972 et leur tabulation a été réalisée par la Société Interactive Market Systems (1). Pour chacun des 57 plans publicitaires, dont le détail figure dans les annexes A, B, C et D de la thèse, nous sommes ainsi en mesure de comparer les estimations données par chaque modèle d'exposition avec la couverture nette et la distribution des occasions de voir observées.

Cependant, on doit noter les limitations suivantes à propos de ces données :

- Le nombre d'insertions dans chaque magazine est limité à deux, afin que les audiences observées puissent être tabulées et comparées avec les estimations des modèles d'exposition.

- Les plans publicitaires ne comprennent pas plus de 4 magazines, de manière que les tabulations de la distribution des occasions de voir ne deviennent pas excessivement incommodes et coûteuses.

A cause de ces deux limitations, la première banque de données est complétée par les résultats de l'enquête Politz de 1958, portant sur 5 magazines (2). Cette enquête fournit la couverture nette de plans publicitaires comprenant

jusqu'à 6 insertions dans chacun des 5 magazines et, par conséquent, elle soumet les modèles d'exposition à un test plus sévère.

Ces deux banques de données, complétées par l'enquête Politz de 1964 portant sur 11 revues (3), utilisée au chapitre 6, constituent, à notre connaissance, les meilleures données disponibles pour une évaluation des modèles d'exposition.

L'approche stochastique et l'approche par simulation ont été utilisées pour estimer la couverture nette et la distribution des occasions de voir, engendrées par des plans publicitaires comportant plusieurs insertions dans plusieurs supports. Nous étudierons d'abord l'approche stochastique.

2 - Les modèles stochastiques.

Ces modèles peuvent être classés selon cinq dimensions en fonction de leurs hypothèses quant aux probabilités d'exposition.

1 - La première hypothèse, présente dans tous les modèles excepté un, concerne l'état stationnaire des probabilités d'exposition d'une parution à une autre et d'un individu à un autre.

2 - La seconde hypothèse concerne l'hétérogénéité de la population. Théoriquement, on peut la représenter par n'importe quelle distribution de probabilité définie sur l'intervalle unité multivariée $[0, 1]$. Cependant, dans les modèles étudiés, le choix a été limité à la densité ponctuelle multivariée, qui présume que tous les individus ont la même probabilité d'exposition à un support donné, bien que cette probabilité diffère d'un support à un autre, et à la densité Bêta multivariée. Les modèles appartenant à la première catégorie sont des modèles à population homogène, ceux qui appartiennent à la seconde catégorie sont des modèles à population hétérogène.

3 - La troisième hypothèse concerne l'hétérogénéité des supports. Théoriquement, on peut concevoir que chaque support soit caractérisé par une distribution de probabilité différente. Toutefois, on fera l'hypothèse que ces distributions appartiennent à une même famille. En outre, si un modèle suppose que tous les supports ont une même probabilité d'exposition moyenne, nous l'appellerons modèle à supports homogènes, tandis que, dans le cas contraire, il sera nommé modèle à supports hétérogènes.

4 - La quatrième hypothèse regarde l'interdépendance des supports. Quand l'exposition de l'individu i au support m est considérée comme n'ayant aucune influence sur les probabilités d'exposition de cet individu à d'autres suppports, nous avons un modèle fondé sur l'indépendance des supports, en d'autres termes, un modèle à duplication aléatoire.

Au contraire, quand la probabilité conditionnelle d'exposition à un support dépend de l'exposition dans un autre support, nous avons un modèle fondé sur l'interdépendance des supports, autrement dit, un modèle à duplication non aléatoire. Dans ce cas, le modèle doit extrapoler l'entière structure de l'interdépendance des supports à partir de la seule connaissance des interdépendances entre paires de supports.

5 - La cinquième hypothèse concerne l'accumulation d'audience. Ici, encore nous aurons deux catégories de modèles. Les modèles d'accumulation aléatoire et les modèles d'accumulation non aléatoire.

Selon cette classification, on peut distinguer 32 familles de modèles d'exposition complets. Toutefois, les modèles non-stationnaires combinant l'hétérogénéité de la population à celle des supports ne se prêtent pas au traitement analytique et ne peuvent être étudiés que par simulation. En outre, pour certains des modèles les plus complexes, on utilisera des approximations plutôt que les distributions exactes, afin de diminuer le poids des calculs. Par conséquent, nous distinguerons les modèles exacts des modèles utilisant des approximations.

La table suivante classe les 14 modèles que nous allons étudier.

21 - <u>Les modèles à population homogène</u>.

211 - <u>Les modèles à duplication aléatoire</u>.

<u>Modèle 1</u> <u>Le modèle binomial</u>

<u>Hypothèses</u> :

a) - La population est homogène. Chaque individu de la population a la même probabilité d'exposition à un support donné.

$$P_{1i} = P_{2i} = \ldots\ldots\ldots P_i \quad \forall i \in M$$

Classification de 14 modèles multi-véhiculaires à parutions multiples

Nom des modèles	La population est	Les supports sont	La duplication est	L'accumulation est	La distribution est
Modèles stochastiques					
1 - Binomial	Homogène	Homogènes	Aléatoire	Aléatoire	Exacte
2 - Binomial multivarié avec indépendance des supports *	Homogène	Hétérogènes	Aléatoire	Aléatoire	Exacte
3 - Multinomial *	Homogène	Hétérogènes	Non-Aléatoire	Aléatoire	Exacte
4 - Binovial Markovien *	Homogène	Homogènes	Non-Aléatoire	Non-Aléatoire	Exacte
5 - Metheringham	Hétérogène	Homogènes	Bêta Binomiale	Bêta Binomiale	Exacte
6 - Bêta matriciel	Hétérogène	Hétérogènes	Non-Aléatoire	Bêta Binomiale	Exacte
7 - Multinomial composé *	Hétérogène	Hétérogènes	Non-Aléatoire	Bêta Binomiale	Exacte
8 - Morgensztern	Hétérogène	Hétérogènes	Non-Aléatoire	Non-Aléatoire	Approximative
9 - Modèle Modal	Hétérogène	Hétérogènes	Non-Aléatoire	Bêta Binomiale	Approximative
10 - Metheringham & M.C. moyen *	Hétérogène	Hétérogènes	Non-Aléatoire	Non-Aléatoire	Approximative
11 - Modèle de casualité constante	Hétérogène	Hétérogènes	Bêta Binomiale	Bêta Binomiale	Approximative
Modèles de simulation					
12 - Friedman	Hétérogène	Hétérogènes	Non-Aléatoire	Non-Aléatoire	Approximative
13 - Probabilités personnelles	Hétérogène	Hétérogènes	Non-Aléatoire	Bêta Binomiale	Approximative
14 - S.C.A.L.	Hétérogène	Hétérogènes	Non-Aléatoire	Non-Aléatoire	Approximative.

* Modèles développés par l'auteur.

b) - Les supports sont homogènes. La probabilité d'exposition est la même pour chaque support.

$$P_i = P_j = \ldots\ldots p \qquad \forall\ i, j \in M$$

c) - La duplication et l'accumulation sont aléatoires.

Duplication alétoire : $P_{ij} = P_i * P_j = p^2 \quad \forall\ i, j \in M, i \neq j$

Accumulation aléatoire : $P_{ii} = P_i * P_i = p^2 \quad \forall\ i \in M$

c) - Les probabilités d'exposition sont stationnaires d'une parution à une autre et d'un individu à un autre.

Ces hypothèses nous permettent de considérer que les insertions dans les différents supports sont équivalentes à des insertions dans un support moyen unique. La probabilité d'exposition à ce support moyen est le rapport de l'audience moyenne de tous les m supports inclus dans le plan publicitaire à la population cible totale.

$$p = \bar{A} / POP \quad \text{avec} \quad \bar{A} = \frac{1}{m} \Sigma_{i=1}^{m} A_i$$

La probabilité de k expositions pour un total de n insertions, quelle que soit leur répartition à travers les divers supports, est donnée par la distribution binomiale :

$$(4) \qquad P(k/p, n) = \binom{n}{k} * p^k * (1 - p)^{n-k}$$

Il en est ainsi à cause de la propriété multiplicative de la distribution binomiale.

<u>Démonstration</u>

Soit n_i le nombre d'insertions effectuées dans le support i, avec $\Sigma_{i=1}^{m} n_i = n$ et soit X_i une variable aléatoire indiquant le nombre d'expositions au support i. X_i suit une distribution binomiale définie par :

$$P(x_i/p, n_i) = \binom{n_i}{k_i} * p^{k_i} * (1 - p)^{n_i - k_i}$$

avec fonction génératrice $(q + ps)^{n_i}$. Ce qui nous intéresse, c'est $K = \Sigma_{i=1}^{m} k_i$, le nombre total d'expositions résultant de n insertions. Définissons donc une nouvelle variable aléatoire S, somme des X_i ; $S = \Sigma_{i=1}^{m} X_i$. S est la convolution de m variables aléatoires indépendantes X_i, distribuées de manière identique,ainsi, la fonction génératrice de S est le produit des fonctions génératrices des X_i.

$$G(S) = (q + ps)^{n_1} (q + ps)^{n_2} \ldots (q + ps)^{n_m}$$

$$G(S) = (q + ps)^{n}$$

En considérant sa fonction génératrice, il apparait que S suit aussi une distribution binomiale de paramètres n et p :

$$P(s = k/p, n) = \binom{n}{k} * p^k * (1 - p)^{n - k}$$

L'espérance du nombre d'individus exposés à exactement k insertions est obtenue en multipliant la probabilité correspondante par le nombre d'individus de la population :

$$E_k^n = P(s = k/p, n) * POP \tag{5}$$

La couverture nette du plan publicitaire est constituée par les individus exposés à au moins une insertion ; ainsi :

$$R_n = P(s \leq 1/p, n) * POP = 1 - P(s = 0/p, n) * POP \tag{6}$$

L'exemple numérique suivant illustre le calcul de la couverture nette et de la distribution des occasions de voir d'un plan publicitaire, suivant les hypothèses du modèle binomial.

Exemple numérique.

Supposons que nous désirions estimer, pour la population adulte des Etats-Unis, la couverture nette et la distribution des occasions de voir produites par deux insertions dans chacun des trois magazines suivants :

"Better Homes & Gardens", "Family Home", et "Ladies' Home Journal".

L'enquête Simmons de 1972, nous fournit leurs audiences en milliers d'individus : B.H.G. = 22 724 ; F.H. = 14 516 ; L.H.J. = 16 634.
L'audience moyenne de ces trois magazines est de $\bar{A}$ = 17 958 et la probabilité moyenne d'exposition est donc égale à p = 17 958 / 130326 = 0, 1 378, étant donné que la population adulte totale est de 130 226 milliers. Le nombre total d'insertions, n, est égal à 6. La distribution des occasions de voir peut être estimée par la formule (5) et comparée avec la distribution observée.

		Estimée	Observée
$E_0^6 = (0{,}8622)^6\ (130\ 326)$	=	53 542	81 087
$E_1^6 = 6\ (0{,}1378)\ (0{,}8622)^5 (130\ 326)$	=	51 342	17 315
$E_2^6 = 15\ (0{,}1378)^2 (0{,}8622)^4 (130\ 326)$	=	20 513	17 426
$E_3^6 = 20\ (0{,}1378)^3 (0{,}8622)^3 (130\ 326)$	=	4 372	6 504
$E_4^6 = 15\ (0{,}1378)^4 (0{,}8622)^2 (130\ 326)$	=	523	4 995
$E_5^6 = 6\ (0{,}1378)^5 (0{,}8622)(130\ 326)$	=	33	1 758
$E_6^6 = (0{,}1378)^6 (130\ 326)$	=	1	1 204

Par la formule (6), la couverture nette estimée du plan est de 76 784 milliers d'individus contre une couverture nette observée de 49 239. Cet exemple, ainsi que les 56 autres inventoriés dans l'appendice E, montre que les résultats du modèle binomial sont extrêmement mauvais. Cela n'a rien de surprenant, puisque ce modèle est basé sur des hypothèses peu réalistes, telles que l'homogénéité des supports et des individus ainsi qu'une duplication et une accumulation au hasard. Cependant, le modèle binomial a été présenté pour les trois raisons suivantes. Tout d'abord, parce que c'est le modèle stochastique le plus simple qui puisse être construit. Il servira ainsi de critère pour mesurer les améliorations obtenues en éliminant certaines hypothèses indésirables. Deuxièmement, parce qu'il a été proposé et utilisé comme tel dans certains modèles de planification des supports tels que ceux de Lee et Burkart (4). Troisièmement, parce qu'il est le moins exigeant quant aux données requises. Il suffit, en effet, de connaître les audiences de chaque support et la taille de la population cible.

Résumé des performances du modèle binomial pour les 57 plans publicitaires

Estimation de la couverture nette

Erreur moyenne absolue sur la couverture nette	=	14 834
Pourcentage moyen d'erreur sur la couverture nette	=	31,61 %
Nombre de surestimations de la couverture nette	=	57
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	0
Erreur type de prévision	=	16 774

Estimation de la distribution des occasions de voir

Chi Deux moyen	=	380,1
Nombre de concordances	=	30

Commentaires

Le modèle d'exposition binomiale surestime systématiquement la couverture nette réelle de 31 % en moyenne. Ce taux de surestimation est quasi-constant d'un plan publicitaire à un autre et il est toujours supérieur à un intervalle de confiance à 95 % autour de la couverture nette réelle. La distribution des occasions de voir est gravement déformée. Le nombre d'individus non exposés est sous-estimé, tandis que le nombre d'individus exposé exactement une fois est surestimé. Les hautes fréquences d'exposition sont aussi systématiquement sous-estimées. Une comparaison des performances du modèle binomial avec les trois autres modèles à population homogène est présentée de la page 453 à la page 457.

Modèle 2 Le modèle binomial multivarié avec indépendances des supports.

Ce modèle est dérivé du modèle binomial en supprimant l'hypothèse de l'homogénéité des supports.

Hypothèses

a) - La population est homogène comme dans le modèle binomial.

b) - Les supports sont hétérogènes: $P_i \neq P_j \quad \forall i, j \in M$

c) - La duplication et l'accumulation sont aléatoires comme dans le modèle binomial.

d) - Les probabilités d'exposition sont stationnaires d'une parution à une autre, comme dans le modèle binomial.

En vertu de ces hypothèses, l'exposition du support est une variable de Bernoulli. La probabilité d'exposition, p_i = A/POP est différente pour chaque support. Afin de faciliter la description du modèle, commençons par le cas bivarié. Considérons deux supports 1 et 2, avec des probabilités d'exposition respectives p_1 et p_2. Supposons que nous ayons effectué n_1 insertions dans le support 1 et n_2 insertions dans le support 2. Dans cette situation, il y a un total de $n_1 * n_2$ résultats possibles qui constituent la matrice d'exposition. Soit E(X, Y) chaque élément de cette matrice d'exposition où x représente le nombre d'expositions au support 1 et y le nombre d'expositions au support 2.

$$X = \{0, 1, 2, \ldots, x, \ldots n_1\}; \; Y = \{0, 1, 2, \ldots, y, \ldots n_2\}$$

D'après notre hypothèse d'indépendance (c), la distribution de probabilité de E (X,Y) est donnée par la distribution binomiale bivariée suivante :

$$P(x = k_1, y = k_2) = P(x = k_1) * P(y = k_2)$$

Puisque les deux variables aléatoires X et Y suivent une distribution binomiale, nous obtenons :

$$(7) \quad P(x = k_1, y = k_2) = \binom{n_1}{k_1}\binom{n_2}{k_2} p_1^{k_1} p_2^{k_2} q_1^{n_1-k_1} q_2^{n_2-k_2} \qquad \begin{matrix} 0 \leq k_1 \leq n_1 \\ 0 \leq k_2 \leq n_2 \end{matrix}$$

L'expression (7) est la probabilité jointe d'exposition aux deux supports.

Toutefois, ce n'est pas la matrice d'exposition en elle-même qui nous intéresse, mais plutôt la couverture nette et la distribution des occasions de voir. Heureusement, on peut les dériver facilement de la matrice d'exposition. D'abord, la probabilité d'au moins une exposition au plan publicitaire est donnée par :

$$P(x \geq 1, \; y \geq 1) = 1 - P(x = 0, y = 0) = 1 - q_1^{n_1} \; q_2^{n_2}$$

Ainsi, la couverture nette du plan publicitaire peut être estimée par :

$$R = (1 - q_1^{n_1} \; q_2^{n_2}) * POP$$

Afin de convertir la distribution jointe des probabilités en distribution des occasions de voir, définissons une nouvelle variable aélatoire S, représentant la somme des expositions au support 1 et au support 2.

$$S = X + Y \qquad 0 \leq S \leq n = n_1 + n_2$$

$$P(s = 0) = P(x = 0, y = 0)$$
$$P(s = 1) = P(x = 0, y = 1) + P(x = 1, y = 0)$$
$$P(s = 2) = P(x = 0, y = 2) + P(x = 1, y = 1) + P(x = 2, y = 0)$$
$$\vdots$$
$$P(s = n) = P(x = n_1, y = n_2)$$

Remarquons que la distribution de probabilité de S n'est pas binomiale.

Le cas bivarié est facilement généralisable au cas multivarié. La probabilité jointe d'exposition à m supports avec des probabilités d'expositions respectives $p_1, p_2, \ldots, p_m$, et un nombre d'insertions $n_1, n_2, \ldots n_m$, est donnée par :

$$(8) \quad P(x_1 = k_1, x_2 = k_2, \ldots x_m = k_m) = \pi_{i=1}^{m} \binom{n_i}{k_i} p_i^{k_i} q_i^{n_i - k_i}$$

La probabilité d'au moins une exposition est simplement :

$$P(x_1 \geq 1, x_2 \geq 1, \ldots, x_m \geq 1) = 1 - \Pi_{i=1}^{m} q_i^{n_i}$$

La distribution des occasions de voir est obtenue comme précédemment en calculant la distribution de probabilité de la somme S des X_i. L'exemple numérique suivant illustre la procédure à suivre pour un plan publicitaire comportant trois magazines.

Exemple numérique

Afin d'évaluer l'effet de la suppression de l'hypothèse de l'homogénéité des supports, calculons la couverture nette et la distribution des occasions de voir du plan publicitaire déjà utilisé pour illustrer le modèle binomial.

		Audiences	Probabilité d'exposition
Population = 130 326 X = B.H.G.	=	22 724	0,1744
Y = F.H.	=	14 516	0,1114
Z = L.H.J.	=	16 634	0,1276

Le plan publicitaire comprend deux insertions dans chaque magazine. En utilisant la formule (8), on obtient la distribution jointe d'exposition aux trois magazines.

Distribution jointe d'exposition		Probabilité	Audiences
$P(x=0,y=0,z=0)$	$= (.8256)^2(.8886)^2(.8724)^2$	= .4096	53 384
$P(x=1,y=0,z=0)$	$= 2(.1744)(.8256)(.8886)(.8724)$	= .1731	22 554
$P(x=2,y=0,z=0)$	$= (.1744)^2(.8886)^2(.8724)^2$	= .0183	2 382
$P(x=0,y=1,z=0)$	$= 2(.8256)^2(.1114)(.8886)(.8724)^2$	= .1027	13 385
$P(x=1,y=1,z=0)$	$= 4(.1744)(.8256)(.1114)(.8886)(.8724)^2$	= .0434	5 655
$P(x=2,y=1,z=0)$	$= 2(.1744)^2(.1114)(.8886)(.8724)^2$	= .0046	597
$P(x=0,y=2,z=0)$	$= (.8256)^2(.1114)^2(.8724)^2$	= .0064	839
$P(x=1,y=2,z=0)$	$= 2(.1744)(.8256)(.1114)^2(.8724)^2$	= .0027	354
$P(x=2,y=2,z=0)$	$= (.1744)^2(.1114)^2(.8724)^2$	= .0003	37
$P(x=0,y=0,z=1)$	$= 2(.8256)^2(.8886)^2(.1276)(.8724)$	= .1198	15 616
$P(x=1,y=0,z=1)$	$= 4(.1744)(.8256)(.8886)^2(.1276)(.8724)$	= .0506	6 594
$P(x=2,y=0,z=1)$	$= 2(.1744)^2(.8886)^2(.1276)(.8724)$	= .0053	697
$P(x=0,y=1,z=1)$	$= 4(.8256)^2(.1114)(.8886)(.1276)(.8724)$	= .0300	3 915
$P(x=1,y=1,z=1)$	$= 8(.1744)(.8256)(.1114)(.8886)(.1276)(.8724)$	= .0127	1 654
$P(x=2,y=1,z=1)$	$= 4(.1744)^2(.1114)(.8886)(.1276)(.8724)$	= .0013	175
$P(x=0,y=2,z=1)$	$= (.8256)^2(.1114)^2(.1276)(.8724)$	= .0019	245
$P(x=1,y=2,z=1)$	$= 4(.1744)(.8256)(.1114)^2(.1276)(.8724)$	= .0008	104
$P(x=2,y=2,z=1)$	$= 2(.1744)^2(.1114)^2(.1276)(.8724)$	= .0001	11
$P(x=0,y=0,z=2)$	$= (.8256)^2(.8886)^2(.1276)^2$	= .0087	1 142
$P(x=1,y=0,z=2)$	$= 2(.1744)(.8256)(.8886)^2(.1276)^2$	= .0037	483
$P(x=2,y=0,z=2)$	$= (.1744)^2(.8886)^2(.1276)^2$	= .0004	51
$P(x=0,y=1,z=2)$	$= 2(.8256)^2(.1114)(.8886)(.1276)^2$	= .0022	286
$P(x=1,y=1,z=2)$	$= 4(.1744)(.8256)(.1114)(.8886)(.1276)^2$	= .0009	121
$P(x=2,y=1,z=2)$	$= 2(.1744)^2(.1114)(.8886)(.1276)^2$	= .0001	13
$P(x=0,y=2,z=2)$	$= (.8256)^2(.1114)^2(.1276)^2$	= .0002	18
$P(x=1,y=2,z=2)$	$= 2(.1744)(.8256)(.1114)^2(.1276)^2$	= .0001	8
$P(x=2,y=2,z=2)$	$= (.1744)^2(.8256)^2(.1276)^2$	= .0000	1

La distribution des occasions de voir est obtenue en additionnant tous les termes de la distribution jointe d'exposition correspondant à un même nombre d'expositions.

		Estimé	Observé
E_0^6	=	53 384	81 087
E_1^6 = 13 385 + 22 554 + 15 616	=	51 555	17 315
E_2^6 = 5 655 + 839 + 2 382 + 3 915 + 6 594 + 1 142	=	20 527	17 465
E_3^6 = 354 + 597 + 245 + 1 654 + 697 + 286 + 483	=	4 316	6 504
E_4^6 = 37 + 104 + 175 + 18 + 121 + 51	=	506	4 995
E_5^6 = 11 + 8 + 13	=	32	1 758
E_6^6	=	1	1 204

Si nous comparons les estimations du modèle binomial multivarié avec celles du modèle binomial, nous constatons qu'elles sont extrêmement proches, fournissant, dans les deux cas, un très médiocre ajustement à la distribution observée. Par conséquent, on peut conclure que la suppression de l'hypothèse de l'homogénéité des supports n'améliore pas l'ajustement à la distribution observée. Comme nous le verrons bientôt, c'est surtout l'hypothèse de duplication et d'accumulation aléatoire qui est responsable des mauvaises performances des modèles binomial et binomial multivarié.

Résumé des performances du modèle binomial multivarié pour 57 plans publicitaires

Estimation de la couverture nette

Erreur moyenne absolue sur la couverture nette	= 15 530
Pourcentage moyen d'erreur sur la couverture nette	= 33,08 %
Nombre de surestimations de la couverture nette	= 57
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	= 0
Erreur type de prévision	= 17 529

Estimation de la distribution des occasions de voir

Chi Deux moyen	= 405,8
Nombre de concordance	= 32

Commentaires

Le modèle binomial multivarié avec indépendance des supports surestime systématiquement la couverture nette réelle de 33 % en moyenne. Cela dépasse de loin l'intervalle de confiance autour de la couverture nette réelle pour tous les 57 plans publicitaires. Le modèle sous-estime la fréquence de non exposition et les fréquences supérieures à trois expositions. D'autre part, il surestime le nombre d'individus exposés à une ou à deux parution. Bien que tenant compte de l'hétérogénéité des supports, le modèle binomial multivarié n'obtient pas de meilleures performances que le modèle binomial homogène. Une comparaison détaillée des performances de ces deux modèles est donnée dans les tables de la page 453 à la page 457.

212 - Les modèles à duplication non aléatoire.

Modèle 3 Le modèle multinomial

Ce modèle est dérivé du modèle précédent en supprimant l'hypothèse de duplication aléatoire.

Hypothèses

a) - La population est homogène comme dans les modèles 1 et 2.
b) - Les supports sont hétérogènes comme dans le modèle 2.
c) - La duplication n'est pas aléatoire, mais l'accumulation est aléatoire.
d) - Les probabilités d'exposition sont stationnaires d'une parution à une autre comme dans les modèles 1 et 2.

Considérons d'abord le cas bivarié, quand le plan publicitaire ne comprend que deux supports, A et B, avec une insertion dans chacun d'eux. Dans cette situation, il y a quatre résultats possibles :

Notation	Description du résultat	Probabilité
U	Exposition à A, Non-exposition à B	p_{10} .... $\}$ p_A
V	Exposition à A, Exposition à B	p_{11} .... $\}$ p_B
W	Non-exposition à A, Exposition à B	p_{01} ....
Z	Non-exposition à A, Non-exposition à B	p_{00}

D'après les hypothèses a et d, les quatre probabilités p_{10}, p_{11}, p_{01}, p_{00}, sont identiques pour tous les individus de la population et elles sont stationnaires d'une parution à la suivante. Ainsi, elles peuvent être estimées directement à partir des audiences de chaque support et de l'audience dupliquée de la paire de supports. Soit A l'audience du support A, B l'audience du support B et A Ω B l'audience dupliquée, nous obtenons :

$$p_A = A/POP \qquad p_B = B/POP \qquad p_{11} = (A \Omega B) / POP$$

$$p_{10} = (A - A \Omega B) / POP \qquad p_{01} = (B - A \Omega B) / POP$$

$$p_{00} = (POP - A - B + A \Omega B) / POP$$

Du moment que p_{11} est différent de $p_A * p_B$, nous tenons compte du fait que les audiences ne sont pas dupliquées au hasard. Toutefois, dans ce modèle, comme dans les deux modèles précédents, nous admettons que l'accumulation est aléatoire, et nous n'utilisons pas les informations contenues dans la couverture nette cumulée à deux parutions de chaque support.

En vertu de ces hypothèses, si un plan publicitaire prévoit le même nombre d'insertions dans chaque support, $n_A = n_B = n$, la probabilité qu'un individu soit exposé k1 fois exclusivement à A, k2 fois exclusivement à B et k3 fois simultanément à B et à A, est donnée par la distribution multinomiale.

$$(9) \quad P(U = k1, W = k2, V = k3) = \frac{n!}{k1!k2!k3!k0!} p_{10}^{k1} p_{01}^{k2} p_{11}^{k3} p_{00}^{k0}$$

avec $n = n_A = n_B = 0, 1, 2, \ldots$

$$k0 = n - k1 - k2 - k3 \qquad k1 + k2 + k3 \leq n$$

$$ki = 0, 1, 2, \ldots n$$

$$p_{10} + p_{01} + p_{11} + p_{00} = 1$$

Puisqu'une somme de variables multinomiales suit une distribution binomiale, on peut noter que la distribution des expositions à chaque support suit une distribution binomiale, de paramètres n, p_A pour le support A et n, p_B pour le support B.

$$P(A = kA) = \binom{n}{kA} p_A^{kA} (1 - p_A)^{n - kA} \text{ avec : } kA = k1 + k3 \; ; \; p_A = p_{10} + p_{11}$$

$$P(B = kB) = \binom{n}{kB} p_B^{kB} (1 - p_B)^{n - kB} \text{ avec : } kB = k2 + k3 \; ; \; p_B = p_{01} + p_{11}$$

Par conséquent, en suivant Kendall et Stuart (5), la distribution multinomiale donnée par (9) peut être transformée en la distribution binomiale bivariée suivante :

$$(10) \quad P(A=kA, B=kB) = \sum_{s=0}^{y} \frac{n!}{s!(kA-s)!(kB-s)!(k0+s)!} p_{11}^{s} p_{10}^{kA-s} p_{01}^{kB-s} p_{00}^{k0+s}$$

avec kA = Nombre d'expositions non exclusives à A = 0, 1, 2, ... n
kB = Nombre d'expositions non exclusives à B = 0, 1, 2, ... n
k0 = n - kA - kB
y = min (kA, kB).

L'expression (9) nous fournit la distribution jointe des expositions, tandis que l'expression (10) nous donne les probabilités de la matrice d'exposition. Toutefois, nous sommes davantage intéressés par la distribution des occasions de voir générée par le plan publicitaire. Soit S le nombre d'expositions au plan publicitaire, nous avons :

$$S = U + W + 2V = A + B$$

La distribution de probabilité de S peut être calculée, soit par l'expression (9), soit par l'expression (10), de la manière suivante :

$$(11) \quad P(S=s) = \sum_{(s=k1+k2-k3)}^{s} P(U = k1, W = k2, V = k3) = \sum_{s=kA+kB}^{s} P(A = kA, B = kB)$$

Cependant, ni l'expression (9) ni l'expression (10) ne peuvent être directement étendues au cas multivarié, car cela suppose la connaissance de toutes les probabilités $p_{ijk...m}$, alors que nous ne connaissons que les probabilités p_{ij} des couples de supports.

Même si nous connaissions les triplications et toutes les duplications d'ordre supérieur à deux, il serait encore difficile de calculer toutes les 2^m probabilités d'exposition nécessaires au modèle multinomial dans le cas de m supports. Mais comme nous ne les connaissons pas, nous devons nous contenter des estimations obtenues à partir des modèles de duplication du chapître 6.

Afin de nous épargner l'estimation de toutes les probabilités d'exposition jointe, nous pouvons modifier légèrement la formulation de notre problème, de telle sorte que nous n'ayons à estimer que les duplications d'ordre supérieur à deux. Ceci s'accomplit en transformant le modèle multinomial en un modèle multinomial univarié décrit par Steyn (6).

Chaque semaine d'un plan publicitaire comprenant m supports avec une insertion dans chaque, peut être considérée comme un évènement ayant m + 1 résultats, respectivement exclusifs qui peuvent être décrits comme suit :

Notation	Description du résultat	Probabilité	Nombre d'expositions
A_0	Aucune exposition à aucun des m supports	p_0	0
A_1	Exposition à chacun des supports exclusivement	p_1	1
A_2	Exposition à chaque couple de supports exclusivement	p_2	2
.	. .	.	
.	. .	.	
.	. .	.	
A_m	Exposition à tous les m supports	p_m	m

La probabilité que n semaines de ce plan publicitaire produisent exactement K expositions est donnée par la distribution multinomiale univariée.

$$(12) \quad P(K/n,p_1, \ldots,p_m) = \Sigma \frac{n!}{k0!k1!\ldots km!} p_0^{k0} p_1^{k1} \ldots p_m^{km}$$

avec K = 0, 1, 2, ... nm

$$P_0 = 1 - p_1 - p_2 - \dots - p_m$$

La somme s'étend sur tous les entiers non négatifs k0, k1, ... km de sorte que $\Sigma_{i=0}^{m} ki = n$ et que

$$\Sigma_{i=1}^{m} i * ki = K$$

Nous ne connaissons pas les probabilités p_0, p_1, ..., p_i, ..., p_m qui représentent les probabilités d'exposition à exactement i supports sur m, mais nous pouvons les calculer grâce au théorème de Waring, comme nous l'avons montré au chapître 6, page

$$p_i = S_i - \binom{i+1}{i} S_{i+1} + \binom{i+2}{i} S_{i+2} \dots \pm \binom{m}{i} S_m$$

où $S_0 = 1$

et S_i = la somme de toutes les probabilités non exclusives d'exposition à i supports.

Par exemple, dans le cas de 3 supports, A, B, C, si nous connaissons p_A, p_N, p_C, les probabilités d'exposition à chaque support, p_{AB}, p_{BC}, p_{AC}, les probabilités d'exposition à chaque paire de supports, et p_{ABC} la probabilité d'exposition à tous les 3 supports, nous pouvons calculer p_0, p_1 et p_2 de la manière suivante :

Données	Probabilités d'exposition non-exclusive		Probabilité d'exposition Exclusive
$p_A = 0{,}20$	$p_{AB} = 0{,}10$	$p_{ABC} = 0{,}08$	$P_0 = 1-S_1+S_2-S_3 = 1-0{,}90+0{,}45-0{,}08 = 0{,}47$
$p_B = 0{,}30$	$p_{BC} = 0{,}20$		$p_1 = S_1-2S_2+3S_3 = 0{,}90-0{,}90+0{,}24 = 0{,}24$
$p_C = 0{,}90$	$p_{AC} = 0{,}15$		$p_2 = S_2-3S_3 = 0{,}45-0{,}24 = 0{,}21$
$S_1 = 0{,}90$	$S_2 = 0{,}45$	$S_3 = 0{,}08$	$p_3 = S_3 = 0{,}08$

Donc, le problème se réduit à estimer S_i pour i > 2 connaissant S_1 et S_2. La meilleure manière de le faire, comme cela a été montré au chapître 6, page 390 , consiste à appliquer la formule d'Hofmans et à calculer ensuite S_i,

grâce aux équations (49) de la page 390 . Alors, la formule multinomiale univariée donne la distribution d'exposition finale. Si le plan publicitaire n'exige pas le même nombre d'insertions dans chaque support, la distribution des occasions de voir est d'abord estimée pour un pseudo-plan où le niveau le plus élevé d'insertions est attribué à tous les supports. Ensuite, le pseudo-plan est ramené au plan réel par la formule de conversion présentée au début de ce chapître.

Les détails de la procédure sont illustrés par l'exemple numérique suivant pour le plan publicitaire n° 34 de la banque de données I.M.S.

Exemple numérique

Magazines	Audiences	Probabilités	Audiences dupliquées		Probabilités	Population
B.H.G.	22 724	0,1744	B.H.G.-F.H.	6 124	0,0470	130 326
F.H.	14 516	0,1114	B.H.G.-L.H.J.	6 578	0,0505	
L.H.J.	16 634	0,1276	F.H.-L.H.J.	4 739	0,0364	
Sommes	53 874	0,4134		17 441	0,1339	

Calculons d'abord la couverture nette générée par une insertion dans chaque support en utilisant la formule d'Hofmans.

Couple de magazines	$(A_i+A_j)A_{ij} \;/\; (A_i+A_j-A_{ij})\; k_{ij}A_{ij}$	$(G.A.)^2 \;/\; (G.A. + kD)$
B.H.G. - F.H.	37 240 * 6 124 / 31 116 = 7 329	
B.H.G. - L.H.J.	39 358 * 6 578 / 32 780 = 7 898	$R_3 = \dfrac{(53\ 874)^2}{53\ 874 + 20\ 810}$
F.H. - L.H.J.	31 150 * 4 739 / 26 441 = 5 583	
	KD = 20 810	R_3 = 38 863

Maintenant calculons S_3 à l'aide de l'équation (49) du chapître 6.

$$S_3 = R_3 - S_1 + S_2 = 38\ 863 - 53\ 874 + 17\ 441 = 2430$$

Ensuite, en utilisant le théorème de Waring, nous pouvons calculer les probabilités d'expositions exclusives.

Distribution des occasions de voir	Probabilités d'expositions exclusives
$E_0^3 = POP - S_1 + S_2 - S_3 = 91\ 463$	$P_0 = 91\ 463 / 130\ 326 = 0{,}7018$
$E_1^3 = S_1 - 2S_2 + 3S_3 = 26\ 282$	$P_1 = 26\ 282 / 130\ 326 = 0{,}2017$
$E_2^3 = S_2 - 3S_3 = 10\ 151$	$P_2 = 10\ 151 / 130\ 326 = 0{,}0779$
$E_3^3 = S_3 = 2\ 430$	$P_3 = 2\ 430 / 130\ 326 = 0{,}0186$

Puisque le plan publicitaire requiert deux insertions dans chaque magazine, la distribution d'exposition à deux insertions s'obtient par l'expression (12) de la manière suivante :

Estimation de la distribution des occasions de voir par modèle multinomial univarié

k_0	k_1	k_2	k_3	K	(2!/k0!k1!k2!k3!)	$p_0^{k_0}$	$p_1^{k_1}$	$p_2^{k_2}$	$p_3^{k_0}$	P(K)	Estimée	Réelle
2	0	0	0	0	1	0,4925	1	1	1	0,4925	64 186	81 087
1	1	0	0	1	2	0,7018	0,2017	1	1	0,2831	36 895	17 315
0	2	0	0	2	1	1	0,0407	1	1	0,1500	19 549	17 465
1	0	1	0	2	2	0,7018	1	0,0779	1			
1	0	0	1	3	2	0,7018	1	1	0,0186	0,0575	7 494	6 504
0	1	1	0	3	2	1	0,2017	0,0779	1			
0	0	2	0	4	1	1	1	0,0061	1	0,0136	1 772	4 995
0	1	0	1	4	2	1	0,2017	1	0,0186			
0	0	1	1	5	2	1	1	0,0779	0,0186	0,0029	378	1 708
0	0	0	2	6	1	1	1	1	0,0004	0,0004	52	1 204

Si nous comparons ces estimations à celles produites par les deux modèles précédents, nous voyons que l'on a obtenu une amélioration en supprimant l'hypothèse de duplication aléatoire. Toutefois, l'ajustement est loin d'être satisfaisant car la couverture nette estimée à 66 140 individus surestime la couverture nette réelle, soit 49 239 individus, de 33 %. Ainsi, on peut conclure que la suppression de l'hypothèse de duplication aléatoire, bien qu'améliorant les résultats, est insuffisante. Pour obtenir un meilleur ajustement, l'hypothèse d'accumulation aléatoire devra aussi être supprimée.

Le tableau suivant résume les performances du modèle multinomial pour 57 plans publicitaires.

Résumé des performances du modèle multinomial pour 57 plans publicitaires

Estimation de la couverture nette

Erreur moyenne absolue sur la couverture nette	=	11 792
Pourcentage moyen d'erreur sur la couverture nette	=	22,20 %
Nombre de surestimations de la couverture nette	=	57
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	0
Erreur type de prévision	=	13 535

Estimation de la distribution des occasions de voir

Chi Deux moyen	=	208,8
Nombre de concordances	=	39

Commentaires

Le modèle multinomial surestime systématiquement la couverture nette réelle de 25 % en moyenne. Il réalise ainsi une meilleure performance que les deux premiers modèles, mais, pour tous les plans publicitaires, il dépasse encore l'intervalle de confiance à 95 % autour de la couverture nette réelle. Le nombre de concordances s'est amélioré, en passant de 30 et 32 pour les deux premiers modèles, à 39 pour le modèle multinomial. Mais le modèle multinomial sous-estime encore tant le nombre d'individus non exposés que les hautes fréquences d'exposition, tandis qu'il surestime le nombre d'individus exposés une fois. Ainsi, on peut conclure qu'il ne suffit pas d'introduire l'hétérogénéité

des supports et une duplication non aléatoire. Un meilleur modèle doit aussi tenir compte de l'accumulation non aléatoire des audiences.

Modèle 4 Le modèle Binomial Markovien (7)

Le modèle 4 est dérivé du modèle 1 en abandonnant les hypothèses de duplication aléatoire et de stationnarité des probabilités d'exposition.

Hypothèses

a) - La population est homogène comme dans le modèle 1.

b) - Les supports sont considérés comme étant homogènes, et ils sont fondus en un support moyen hypothétique.

c) - La duplication et l'accumulation ne sont pas aléatoires.

$$p_{ij} \neq p_i * p_j \qquad \forall\, i, j \in M$$

$$p_{ii} \neq p_i^2 \qquad \forall\, i \in M$$

d) - La probabilité d'exposition dépend de l'exposition à la parution précédente.

Ces hypothèses nous permettent de remplacer un ensemble de supports par un support unique, dans lequel nous effectuons un nombre d'insertions égal au nombre total d'insertions du plan publicitaire. Ainsi, ce modèle réduit le problème de l'exposition à plusieurs supports à un problème d'accumulation. Cependant, parmi les modèles stochastiques d'accumulation présentés au chapitre 5, il n'y en qu'un d'adéquat. Puisque nous rejetons l'hypothèse d'une duplication et d'une accumulation aléatoire, le modèle binomial d'accumulation est éliminé. Puisque nous admettons que la population est constituée d'individus homogènes, les modèles bêta-binomial et Gamma-Poisson sont aussi écartés. Seul le modèle binomial Markovien est adéquat. Ce modèle combine une audience moyenne constante avec la non stationnarité des probabilités individuelles d'exposition. Ces caractéristiques sont obtenues par la spécification d'une matrice de probabilités de transition qui rendent l'exposition à la parution suivante (ou à l'insertion suivante) dépendante de l'exposition (ou de la non-exposition) à la parution précédente (ou insertion).

Considérons d'abord la matrice des probabilités de transition qui correspond au modèle binomial simple. Soit $X_n = 0$ la non exposition, et $X_n = 1$ l'exposition à la nième insertion. Par définition, l'indépendance, la probabilité d'exposition à la nième insertion ne dépend pas de l'exposition

à la n — 1 ième insertion. Autrement dit, la probabilité conditionnelle d'exposition est égale à la probabilité inconditionnelle :

$$p(X_n = 1/X_{n-1} = 1) = p(X_n = 1/X_{n-1} = 0) = p$$

$$p(X_n = 0/X_{n-1} = 1) = p(X_n = 0/X_{n-1} = 0) = 1 - p = q$$

La probabilité d'exposition inconditionnelle, p, est défini exactement comme dans le modèle binomial simple :

$$p = \bar{A} / POP \qquad \text{où } \bar{A} = \frac{1}{m} \Sigma_{i=1}^{m} A_i$$

Ainsi, la matrice de transition pour le modèle binomial simple est la suivante :

$$\begin{array}{cc} & \begin{array}{cc} X_n = 0 & X_n = 1 \end{array} \\ \begin{array}{c} X_{n-1} = 0 \\ X_{n-1} = 1 \end{array} & \begin{bmatrix} q & p \\ q & p \end{bmatrix} \end{array}$$

Cependant, nous savons que les duplications observées ainsi que les accumulations sont différentes de celles qui résultent de l'hypothèse d'indépendance. Ainsi, la probabilité d'exposition à l'insertion suivante doit être supérieure à p, quand l'individu a été exposé et inférieure à p, quand l'individu n'a pas été exposé. En posant :

$$P(X_n = 1/X_{n-1} = 1) = \lambda \qquad \text{avec} \qquad p < \lambda < 1 \ ;$$

Nous pouvons faire varier le degré de duplication et d'accumulation existant entre deux insertions. Si les deux insertions sont dans deux supports différents, λ est un indice de duplication. Si les deux insertions sont dans deux parutions différentes d'un même support, λ est un indice d'accumulation.

Dans le cas de notre support moyen hypothétique, λ sera estimé par le rapport du nombre de doubles expositions au nombre d'expositions. En termes de probabilités, λ est le rapport de la probabilité de double exposition à la probabilité d'exposition.

$$(13) \qquad \lambda = P(X_n = 1/X_{n-1} = 1) = \frac{P(X_n = 1,\ X_{n-1} = 1)}{P(X_n = 1)}$$

Dans cette équation, le dénominateur est égal à la probabilité inconditionnelle d'exposition.

$$P(X_n = 1) = P(X_{n-1} = 1) = P(X_{.} = 1) = p = \overline{A}/POP$$

Le numérateur est la probabilité de deux expositions successives, ou probabilité de double exposition. Il peut être exprimé de la manière suivante :

$$P(X_n = 1,\ X_{n-1} = 1) = \frac{\dfrac{(\text{Somme des duplications}+\text{Somme des accumulations})}{\text{Nombre de doubles-expositions}}}{POP}$$

Le nombre de doubles-expositions est naturellement fonction du nombre de supports inclus dans le plan publicitaire et du nombre d'insertions dans chaque support. Si nous représentons par n le nombre d'insertions dans chaque support et par m le nombre de supports, le nombre de doubles expositions est de : $\binom{nm}{2} = nm\,(nm - 1)/2$.

Par exemple, un plan publicitaire comprenant trois supports avec deux insertions dans chacun d'eux fournit un total de 15 occasions de double exposition, dont 12 provenant de duplications et 3 d'accumulations.

En général, pour un plan publicitaire comprenant m supports avec n insertions dans chacun d'eux, il y aura $\binom{n}{2} * m$ doubles expositions dues à des accumulations et $\binom{m}{2} * n^2$ doubles expositions dues à des duplications.

Soit A_{ij} l'audience dupliquée entre les supports i et j, quelle que soit la parution, et soit $A_i^{k,l}$, l'audience commune à deux parutions, k et l du même support i, la valeur de la probabilité de double exposition peut être estimée par

$$(14) \quad P(X_n = 1,\ X_{n-1} = 1) = \frac{\left[\left(\Sigma_{i=1}^{m}\ \Sigma_{k=1}^{n}\ \Sigma_{l=k+1}^{n-1} A_i^{k,l}\right) + n^2 * \left(\Sigma_{i=1}^{m}\ \Sigma_{j=i+1}^{m-1} A_{ij}\right)\right] / \binom{nm}{2}}{POP}$$

Finalement, en combinant les équations (13) et (14), λ peut être estimée par :

$$(15) \quad \lambda = \frac{\left[\left(\Sigma_{i=1}^{m} \; \Sigma_{k=1}^{n} \; \Sigma_{l=k+1}^{n-1} \; A_i^{k,l}\right) + n^2\left(\Sigma_{i=1}^{m} \; \Sigma_{j=i+1}^{m=1} A_{ij}\right)\right] * m}{\binom{nm}{2} * \; \Sigma_{i=1}^{m} A_i}$$

Mais, puisque $m/\binom{nm}{2} = 2/n(nm - 1)$, nous obtenons ainsi :

$$(16) \quad \hat{\lambda} = \frac{2 * \left[\left(\Sigma_{i=1}^{m} \; \Sigma_{k=1}^{n} \; \Sigma_{l=k+1}^{n-1} \; A_i^{k,l}\right) + n^2 \left(\Sigma_{i=1}^{m} \; \Sigma_{j=i+1}^{m-1} A_{ij}\right)\right]}{n(nm - 1) \; \Sigma_{i=1}^{m} A_i}$$

L'équation (16) nous permet d'estimer la valeur de λ. Nous pouvons calculer maintenant les autres termes de la matrice de transition. Nous avons :

$$p(X_n = 1/X_{n-1} = 1) + (P(X_n = 0/X_{n-1} = 1) = 1.$$

Nous obtenons :

$$P(X_n = 0/X_{n-1} = 1) = 1 - \lambda.$$

En vertu de l'hypothèse de la constance de l'audience d'une parution, la proportion d'individus exposés reste identique d'une parution à l'autre. Cela peut se traduire par les deux équations suivantes :

$$\begin{cases} P(X_n=1/X_{n-1}=1) * P(X_{n-1}=1) + P(X_n=1/X_{n-1} = 0) * P(X_{n-1} = 0) = P(X=1)=p \\ P(X_n=0/X_{n-1}=1) * P(X_{n-1} = 1) + P(X_n=0/X_{n-1}=0) * P(X_{n-1}=0)=P(X=0)=q \end{cases}$$

En résolvant ces deux équations pour les probabilités de transition inconnues, nous obtenons :

$$\begin{cases} P(X_n = 1/X_{n-1} = 0) = (1 - \lambda) \; p/q \\ P(X_n = 0/X_{n-1} = 0) = (1 - 2 \; p + \lambda p) \; / \; q \end{cases}$$

Le modèle binomial Markovien est ainsi entièrement spécifié par la matrice des probabilités de transition suivante :

$$(17) \qquad \begin{array}{c|cc} & X_n = 0 & X_n = 1 \\ \hline X_{n-1} = 0 & (1 - 2p + \lambda p)/q & (1 - \lambda)\ p/q \\ X_{n-1} = 1 & (1 - \lambda) & \lambda \end{array}$$

On peut noter que quand $\lambda = p$, la matrice des probabilités de transition se réduit à la matrice du modèle binomial simple. La distribution jointe d'exposition du modèle est extrêmement complexe, au point d'être presque inutilisable.

$$(18) \quad P(X_1 = x_1, x_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1) P(X_2 = x_2 / X_1 = x_1) \ldots$$

$$\ldots * P(X_n = x_n / X_{n-1} = x_{n-1})$$

avec $x_i = 0, 1$, et $i = 1, 2, \ldots, n$

$$= \lambda^r (1-\lambda)^{2(s-r)-t} (1-2p+\lambda p)^{n-1-2s+t+r} p^{s-r} q^{-(n-2-s+t)}$$

avec $r = \Sigma_{i=2}^{n} x_{i-1} x_i$; $s = \Sigma_{i=1}^{n} x_i$; $t = x_1 + x_n$

D'après cette équation, nous voyons que la probabilité de n'importe quelle suite d'exposition/non-exposition ne dépend que de λ et de p, les paramètres du modèle, et de r, s, t, respectivement : la somme des produits successifs, le nombre total d'expositions et la somme de la première et de la dernière exposition. Les paramètres r, s, t, caractérisent la suite particulière d'expositions dont la probabilité est calculée.

Pour les deux cas-limites : aucune exposition et exposition à chaque insertion ; la formule (18) se simplifie considérablement. Dans le cas d'absence d'exposition, nous avons r = 0, s = 0 et t= 0. Ainsi, la formule (18) se réduit à :

$$(19) \quad P(X_1 = 0, X_2 = 0, \ldots, X_n = 0) = (1 - 2p + \lambda p)^{n-1} * q^{2-n}$$

Dans le cas d'exposition à chaque insertion, nous avons r = n - 1, s = n, et t = 2. Ainsi, (18) se réduit à :

$$(20) \quad P(X_1 = 1, X_2 = 1, \ldots, X_n = 1) = \lambda^{n-1} * p$$

De l'équation (19), nous pouvons dériver la couverture nette d'un plan publici-

taire comprenant n insertions dans m supports :

$$(21) \qquad R_n = \left[1 - (1 - 2p + \lambda p)^{n-1} * q^{2-n}\right] * POP$$

L'estimation de la distribution des occasions de voir présente un problème plus difficile. Il serait trop incommode de calculer la somme du nombre d'expositions, $S = \Sigma_{i=1}^{n} x_i$, directement à partir de l'équation (18) parce que les x_i ne sont pas des variables aléatoires indépendantes. Heureusement, nous n'avons pas besoin de connaître une expression générale pour la distribution de S pour en calculer les valeurs numériques. Grâce à la rapidité de l'ordinateur, il est possible de calculer par récurrence les probabilités de 2^n suites différentes d'exposition/non-exposition générées par n insertions, à partir de la connaissance des probabilités des 2^{n-1} suites générées par n-1 insertions. La procédure est illustrée par l'exemple numérique suivant.

Exemple numérique

Nous appliquons le modèle binomial Markovien à la même série de supports utilisés pour les trois modèles précédents. Les données nécessaires sont les audiences de chaque support, les duplications par couple de supports et la population cible totale.

Supports	Audiences 1 parution	Audiences 2 parutions	Audiences dupliquées	Population
B.H.G.	22 724	30 249	B.H.G.-F.H. = 6 124	
F.H.	14 516	20 293	B.H.G.-L.H.J= 6 578	130 326
L.H.J.	16 634	22 598	F.H.-L.H.J. = 4 739	
Audiences brutes	53 874	73 140	17 441	

Puisque les revues sont considérées homogènes, la probabilité d'exposition à la revue composite est estimée comme dans le modèle binomial par :

$$\hat{p} = \bar{A} / POP = 53\,874 / 3 * 130\,326 = 0{,}13779$$

Nous devons calculer ensuite la somme des audiences dupliquées et accumulées. La somme des audiences dupliquées est de 17 441. La somme des audiences accumulées s'obtient très facilement en soustrayant l'audiences cumulative brute du double des audiences brutes.

Audience accumulée = 107 748 - 73 140 = 34 608

Il y a deux insertions dans chaque support et 3 supports. Ainsi n = 2, et

m = 3, et, d'après l'équation (16) :

$$\hat{\lambda} = \frac{2 * [34\ 608 + (2)^2 * 17\ 441]}{2*5*53\ 874} = \frac{104\ 372}{269\ 370} = 0{,}3875$$

Le schéma suivant permet de visualiser les étapes nécessaires à l'estimation du paramètre λ.

Schéma d'imbrication des audiences dupliquées et des audiences cumulées

		Première parution			Deuxième parution		
		BHG_1	FH_1	LHJ_1	BHG_2	FH_2	LHJ_2
Première parution	BHG_1	22 724	6 124	6 578	15 199	6 124	6 578
	FH_1		14 516	4 739	6 124	8 739	4 730
	LHJ_1			16 874	6 578	4 730	10 670
Deuxième Parution	BHG_2				22 724	6 124	6 578
	FH_2					14 516	4 730
	LHJ_2						16 874

L'audience dupliquée est soulignée ---------

L'audience accumulée est soulignée ————

D'après ce schéma, on peut facilement constater que la double exposition a lieu 15 fois, 12 fois entre différents supports (duplication) et 3 fois entre différentes parutions d'un même support (accumulation). L'audience moyenne de double exposition est de :

$$[4 * (6\ 124 + 6\ 578 + 4\ 739) + (15\ 199 + 8\ 739 + 10\ 690)] \ / \ 15 = 6\ 958$$

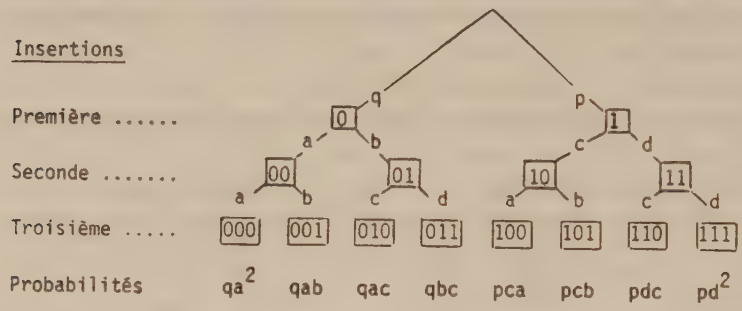
Ainsi, nous pouvons estime λ par : $\hat{\lambda} = 6\ 958 / 17\ 950 = 0{,}3875$

λ étant estimé, il est alors possible de calculer les 3 autres paramètres de la matrice de transition. Selon l'équation (17), nous avons :

$$a = (1 - 2p + \lambda p)/q = 0{,}9021 \ ; \quad b = (1 - \lambda)\ p/q = 0{,}\ 0979$$

$$c = 1 - \lambda = 0{,}6125$$

Maintenant, nous pouvons estimer la couverture nette et la distribution des occasions de voir générées par le plan publicitaire. Le diagramme ramifié suivant indique la génération des probabilités d'exposition à chaque nouvelle insertion. Les lettres sur les branches sont les probabilités et les nombres dans les cases sont les suites d'exposition résultantes.



L'examen de ce diagramme fait apparaître que le nombre de cases double à chaque nouvelle insertion. Il apparaît aussi que les probabilités de nième rang s'obtiennent en multipliant celles du rang précédent selon la règle suivante : les cases dont le chiffre est pair voient leurs possibilités multipliées par a ou b, tandis que celles dont le chiffre est impair voient leurs probabilités multipliées par c ou d.

En termes mathématiques, si nous appelons n les rangs, j le numéro de la case et P(n, j) la probabilité correspondant à la jième case du nième rang, nous avons la relation de récurrence suivante :

$$(22)\quad \left.\begin{aligned} P(n,\ 2j-1) &= P(n-1,j) * a \\ P(n,\ 2j) &= P(n-1,j) * b \end{aligned}\right\} \text{si J est pair}$$

$$\left.\begin{aligned} P(n,2j-1) &= P(n-1,j) * c \\ P(n,2j) &= P(n-1,j) * d \end{aligned}\right\} \text{si j est impair}$$

En utilisant ces relations et en additionnant les termes de la distribution jointe d'exposition qui aboutissent au même nombre d'expositions, nous obtenons la distribution des occasions de voir, comme cela est indiqué dans le tableau suivant.

Modèle Binomial Markovien

Estimation de la distribution des occasions de voir

Expositions	INSERTIONS 1ère	2ème	3ème	4ème	5ème	6ème	Distribution des occasions de voir Estimée	Observée
0	0,8622	0,7690	0,6859	0,6131	0,5455	0,5151	67 133	81 087
1	0,1378	0,1864	0,2292	0,2612	0,2825	0,2636	34 356	17 315
2		0,0446	0,0704	0,0970	0,1212	0,1361	17 736	17 465
3			0,0144	0,0261	0,0402	0,0579	7 545	6 504
4				0,0047	0,0094	0,0204	2 653	4 995
5					0,0015	0,0057	746	1 758
6						0,0012	157	1 204

Si nous comparons ces estimations avec celles des Modèles 1 et 2, nous observons un progrès sensible vers un meilleur ajustement. Les résultats sont aussi supérieurs à ceux du Modèle 3. Mais, l'ajustement n'est pas encore satisfaisant. Le Modèle Binomial Markovien surestime la couverture nette réelle de 28 % contre 33 % pour le Modèle Multinomial, 56 % pour le Modèle Binomial multivarié et 55 % pour le Modèle Binomial simple.

On peut ainsi conclure que l'abandon des hypothèses de duplication aléatoire et d'accumulation aléatoire et leur remplacement par l'hypothèse de non stationnarité ne conduit pas à un modèle satisfaisant. Afin d'obtenir un meilleur ajustement, il faut abandonner l'hypothèse de non stationnarité et introduire l'hétérogénéité de la population.

La table suivante résume les performances du Modèle Binomial Markovien pour nos 57 plans publicitaires usuels.

Résumé des performances du Modèle Binomial Markovien pour 57 plans publicitaires

Estimation de la couverture nette

Erreur moyenne absolue sur la couverture nette	=	10 161
Pourcentage moyen d'erreur sur la couverture nette	=	22,18 %
Nombre de surestimations de la couverture nette	=	57
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	0
Erreur type de prévision	=	11 597

Estimation de la distribution des occasions de voir

Chi Deux moyen	=	166,3
Nombre de concordances	=	42

Commentaires

Le Modèle Binomial Markovien surestime systématiquement la couverture nette réelle de 22 % en moyenne. C'est le meilleur résultat parmi les 4 modèles à population homogène. Cependant, la qualité des estimations reste médiocre, puisque la surestimation excède les bornes d'un intervalle de confiance à 95 % autour de la couverture nette réelle pour chacun des 57 plans publicitaires. Comme pour les 3 modèles précédents, le Modèle Binomial Markovien sous estime la fréquence de non exposition ainsi que les fréquences supérieures à 3 expositions, alors qu'il surestime les fréquences d'une, de deux et de trois expositions.

Il est évident qu'aucun des modèles à population homogène n'est satisfaisant. Il faut donc élaborer des modèles plus complexes qui tiennent compte de l'hétérogénéité des individus. Avant d'étudier ces modèles dans la section suivante de ce chapître, nous présentons deux tableaux décrivant en détail les performances relatives de chacun des modèles homogènes, plan publicitaire par plan publicitaire. Le premier tableau fournit la couverture nette estimée et le pourcentage d'erreur, le second tableau, les valeurs du test Chi Deux de l'ajustement de la distribution estimée des occasions de voir.

Couverture nette comparée des 4 modèles d'exposition à population homogène

Plan publicitaire	Couverture nette observée	Couverture nette estimée: Binomial	% E	Binomial indp.	% E	Multinomial	% E	Bino. Markovien	% E
1	95,511	117,743	23.3	118,077	23.6	112,590	17.9	111,486	16.7
2	96,644	118,343	22.5	118,599	22.7	114,216	18.2	112,734	16.6
3	94,547	118,122	24.9	118,383	25.2	114,609	21.2	112,444	18.9
4	64,543	85,296	32.2	85,298	32.2	80,033	25.2	78,215	21.2
5	57,047	79,986	40.2	79,843	39.9	80,830	41.7	76,525	34.1
6	51,838	71,829	38.6	71,874	38.6	71,646	38.2	68,472	32.1
7	31,645	45,956	45.2	46,014	45.4	44,068	39.2	41,947	32.5
8	28,843	43,458	50.7	43,469	50.7	39,645	37.4	38,718	34.2
9	30,800	44,138	43.3	44,200	43.5	41,664	35.3	40,206	30.5
10	11,527	20,304	76.1	19,792	71.7	15,931	38.2	16,300	41.4
11	16,523	24,800	50.1	24,838	50.9	22,621	36.9	22,148	34.0
12	15,425	21,490	39.3	21,518	39.5	20,376	32.1	19,670	27.5
13	66,141	89,268	34.9	92,435	39.7	87,180	31.8	83,596	26.4
14	62,399	82,572	32.3	85,031	36.3	80,768	29.4	77,612	24.4
15	65,675	87,147	32.7	90,509	37.8	85,009	29.4	81,674	24.4
16	26,106	38,318	46.8	38,574	47.7	35,411	35.6	34,660	32.8
17	23,980	31,698	32.2	32,076	33.8	30,939	29.0	29,628	23.6
18	14,548	93,604	62.2	23,639	62.4	19,821	36.2	20,134	38.4
19	13,598	21,506	58.2	21,543	58.4	18,941	39.3	18,746	37.8
20	56,531	74,473	31.7	77,120	36.4	69,711	23.3	68,681	21.5
21	50,095	58,899	17.6	62,890	25.5	61,270	22.3	57,010	13.8
22	51,426	77,807	51.3	78,392	52.4	69,416	34.9	69,138	34.4
23	60,542	85,305	40.9	85,483	41.2	78,688	29.9	77,628	28.2
24	37,766	51,260	35.7	52,245	38.3	51,102	35.3	46,780	23.9
25	82,880	105,568	27.4	107,282	29.4	104,953	26.6	100,665	21.5

Couverture nette comparée des 4 modèles d'exposition à population homogène

(suite)

Plan publicitaire	Couverture nette observée	Couverture nette estimée: Binomial	% E	Binomial indp.	% E	Multinomial	% E	Bino. Markovien	% E
26	84,402	106,994	26.8	107,672	27.6	105,200	24.6	101,598	20.4
27	92,627	116,278	25.5	116,808	26.1	112,751	21.7	110,430	19.2
28	77,643	97,506	25.6	99,264	27.8	95,805	23.4	92,658	19.3
29	73,215	97,445	33.1	99,673	36.1	93,225	27.3	90,909	24.2
30	50,600	73,127	44.5	73,209	44.7	66,644	31.7	65,986	30.4
31	9,349	14,306	53.0	14,313	53.1	11,562	23.7	11,373	21.6
32	13,731	21,589	57.2	21,654	57.7	18,582	35.3	18,684	36.1
33	21,487	38,892	81.0	38,920	81.1	30,122	40.2	31,538	46.8
34	49,239	76,783	55.9	76,939	56.2	66,132	34.3	63,193	28.3
35	50,460	77,747	54.1	77,891	54.4	66,838	32.4	63,848	26.5
36	58,819	75,918	29.1	77,862	32.4	73,326	24.7	69,532	18.2
37	22,693	33,339	46.9	33,350	46.9	31,892	40.5	29,282	29.0
38	89,625	112,697	25.7	112,807	25.9	109,441	22.1	104,632	16.7
39	41,935	57,515	37.2	57,977	38.2	57,541	37.2	53,832	28.4
40	67,547	89,043	31.8	89,713	32.8	84,067	24.5	80,637	19.4
41	44,490	64,810	45.7	62,205	46.6	56,940	27.9	57,877	30.1
42	32,436	52,228	61.0	52,663	62.3	43,712	34.6	42,721	31.7
43	39,172	56,660	44.6	57,245	46.1	52,508	34.0	49,471	26.3
44	50,077	60,142	20.1	63,499	26.8	60,972	21.7	56,646	13.1
45	74,913	92,129	23.0	95,186	27.1	91,449	22.1	88,215	17.7
46	49,047	58,865	20.0	59,614	21.5	57,769	17.8	56,478	15.1
47	38,832	56,493	45.5	56,508	45.5	48,402	24.6	49,637	27.8
48	17,302	27,846	60.9	28,167	62.8	24,258	40.2	24,145	39.6
49	24,097	27,759	15.2	28,058	16.4	27,516	14.2	26,907	11.7
50	20,238	27,516	36.0	27,521	36.0	24,469	20.9	24,921	23.1

Couverture nette comparée des 4 modèles d'exposition à population homogène

(suite)

Plan publicitaire	Couverture nette observée	Couverture nette estimée: Binomial	% E	Binomial indp.	% E	Multinomial	% E	Bino. Markovien	% E
51	10,747	16,390	52.5	16,627	54.7	14,702	36.8	14,541	35.3
52	10,858	12,533	15.4	12,690	16.9	12,415	14.3	12,157	12.0
53	9,519	12,599	32.3	12,610	32.5	11,313	18.8	11,518	21.0
54	7,242	9,774	34.9	9,905	36.8	9,028	24.7	8,959	23.7
55	3,114	3,539	13.6	3,548	13.9	3,496	12.7	3,445	10.6
56	616	1,039	68.7	1,039	68.7	822	33.4	858	39.3
57	2,638	3,178	20.5	3,242	22.9	3,104	17.7	3,051	15.7
Pourcentage moyen d'erreur			38.01%		33.08%		23.52%		22.18%

Distribution comparée des occasions de voir de 4 modèles à population homogène

Valeur du Chi Deux

Plan Publicitaire	Binomial Simple	Binomial Indépendant	Multinomial	Binomial Markovien
1	279.8	290.6	133.6	133.3
2	295.5	303.3	170.4	163.2
3	375.4	384.0	269.8	228.9
4	309.7	309.8	184.6	149.2
5	391.3	394.5	432.7	291.1
6	323.6	325.3	316.4	224.6
7	364.5	367.9	275.7	197.0
8	350.5	351.1	185.5	157.7
9	300.6	303.5	199.4	152.5
10	223.7	224.2	49.1	61.3
11	189.8	191.8	95.6	82.5
12	128.1	129.3	84.8	64.2
13	494.2	612.2	425.5	322.9
14	349.2	428.8	297.3	219.4
15	379.1	491.8	319.3	234.9
16	251.4	263.3	138.9	117.6
17	134.9	148.6	109.9	74.0
18	223.8	225.9	63.1	74.1
19	209.2	211.5	84.7	81.1
20	231.4	305.9	126.1	110.0
21	98.9	186.6	146.9	71.9
22	602.1	629.4	270.7	281.2
23	447.9	454.1	255.0	226.6
24	296.5	338.5	290.0	151.4
25	405.9	450.3	387.4	293.2
26	383.7	400.3	337.6	265.6
27	374.8	390.5	275.9	235.1
28	262.3	303.0	226.8	166.4
29	392.4	459.9	277.1	229.9
30	444.6	447.7	227.1	213.9
31	137.7	138.1	27.6	25.6
32	185.3	189.0	60.2	65.0
33	572.3	574.6	111.6	164.3
34	647.2	655.1	224.9	168.3
35	569.7	576.2	191.6	139.4
36	222.8	275.7	160.5	92.9
37	280.8	281.4	207.5	111.4
38	380.2	383.0	290.7	203.9
39	365.4	385.6	369.1	230.8
40	305.2	324.0	183.7	122.2

Distribution comparée des occasions de voir de 4 modèles à population homogène

(suite)

Valeur du Chi Deux

Plan Publicitaire	Binomial simple	Binomial Indépendant	Multinomial	Binomial Markovien
41	348.2	361.5	121.0	147.9
42	485.3	509.5	144.4	124.3
43	371.7	397.9	205.2	132.8
44	114.9	190.0	130.8	68.1
45	207.9	276.7	190.9	136.3
46	220.6	256.4	165.9	132.9
47	613.9	615.0	154.2	226.9
48	601.1	641.9	246.4	247.2
49	179.8	215.8	150.1	107.9
50	551.3	552.2	241.5	218.5
51	616.8	671.7	301.6	285.1
52	199.1	246.8	162.9	119.5
53	612.4	618.0	213.1	243.3
54	348.2	384.8	179.4	166.7
55	224.2	234.7	174.3	153.6
56	369.2	369.3	69.5	102.4
57	362.2	421.9	306.2	269.2
Chi Deux moyen	380.1	405.8	208.8	166.3

22 - Les Modèles à Population Hétérogène.

221 - Les Modèles à Distribution Exacte.

Modèle 5 Le Modèle de Metheringham ou modèle Bêta Binomial Multivarié avec homogénéité des supports.

La manière la plus simple d'étendre le modèle d'accumulation Bêta Binomial au cas de plusieurs supports consiste à admettre que la distribution des probabilités d'exposition qui caractérise l'hétérogénéité de la population est identique pour chaque support. Cette hypothèse, certes fort peu réaliste, nous permettra d'étudier en détail le modèle Bêta Binomial Multivarié. Cette étude ouvrira la voie à la construction de modèles plus réalistes, mais aussi plus complexes.

Hypothèses

a) - La population est hétérogène. Chaque individu est caractérisé par sa propre probabilité personnelle d'exposition. Ces probabilités suivent une distribution Bêta sur l'ensemble de la population.

b) - Les supports sont homogènes. Cela ne signifie pas qu'un individu donné ait nécessairement la même probabilité d'exposition à chaque support. Cela signifie seulement que les probabilités individuelles sont distribuées de telle sorte que tous les supports sont caractérisés par la même distribution Bêta.

c) - Les supports étant supposés homogènes, ils sont fondus en un support moyen unique. Le modèle remplace les accumulations de chaque support par leur moyenne, les duplications par leur moyenne et, de plus, l'accumulation moyenne et la duplication moyenne sont à leur tour remplacées par leur moyenne.

d) - Les probabilités d'exposition sont stationnaires d'une parution à l'autre.

En vertu de ces hypothèses, la probabilité d'obtenir le vecteur d'exposition $K = \{k1,k2,...,ki,...,km\}$

étant donné le vecteur d'insertion S = {n1,n2,...,ni,...,nm}, et compte tenu des deux paramètres A et B qui décrivent l'hétérogénéité de la population, est donnée par la Distribution Bêta Binomiale Multivariée (DBBM) :

$$(23) \qquad P(K/A,B,S) = \frac{\beta\,\{A + \Sigma_{i=1}^{m}\ ki, B + \Sigma_{i=1}^{m}\ (ni-ki)\}}{\beta\,(A,B)\pi_{i+1}^{m}\ \{(ni+1)\ \beta\ (ki+1,\ ni-ki+1)\}}$$

$$\text{avec :}\quad \left.\begin{array}{l} i = 1,2,\ldots,m \\ ni = 0,1,2,\ \ldots \\ ki = 0,1,2,\ldots \end{array}\right\} \quad \text{entiers} \qquad \begin{array}{l} A,\ B > 0,\ \text{réel} \\ 0\ \leq ki \leq ni \end{array}$$

Cette distribution a été étudiée pour la première fois, dans le cas bivarié et trivarié, par Ishii et Hayakawa (8) en 1960, et elle a aussi paru dans un récent article de Chatfield et Goodhart (9). Quand pour la première fois Metheringham décrivit son modèlé en 1964 (10), il n'était pas au courant de l'ouvrage des deux statisticiens Japonais.

En fait, il nia même que sa formule, dérivée empiriquement, eût un lien quelconque avec la distribution Bêta ou avec toute autre distribution de probabilité. Greene fut le premier à souligner l'identité (11) et nous avons montré, au chapitre 6, que la formule de Metheringham n'est autre que l'estimation par la méthode des moments des paramètres de la distribution Bêta des probabilités d'expositions à un support moyen hypothétique. La distribution Bêta Binomiale Multivariée possède plusieurs propriétés intéressantes que nous allons décrire dans le cas bivarié afin de simplifier la notation.

- Les distributions marginales sont Bêta Binomiales avec les paramètres A, B, ni ; ni étant le nombre d'insertions dans le support i dont on considère la distribution marginale. Leur corrélation est donnée par :

$$(24) \qquad \rho_{i,j} = \left[\frac{ninj}{(ni+A+B)(nj+A+B)}\right]^{1/2}$$

- La distribution conditionnelle d'exposition au support j, étant donné ki expositions au support i, est aussi une distribution Bêta Binomiale.

$$(25) \qquad P(kj/ki) = \frac{\beta(A+ki+kj, ni+nj+B-ki-kj)}{(nj+1)\beta(A+ki, ni+B-ki)\beta(kj+1, nj+1-kj)}$$

La moyenne et la variance de la distribution conditionnelle sont données par :

$$(26) \qquad E(kj/ki) = nj(A+ki) \,/\, (ni+A+B)$$

$$(27) \qquad V(kj/ki) = nj(A+ki)(ni+B-ki)(ni+nj+A+B)/(ni+A+B)^2\,(ni+A+B+1)$$

L'expression (26) est particulièrement intéressante, puisqu'elle indique que la moyenne des kj est une fonction linéaire du nombre d'expositions à l'autre support. En effet, comme nous le verrons, l'un des meilleurs modèles d'approximation tire profit de cette caractéristique pour simplifier l'agrégation des supports.

- Enfin, la distribution de la somme du nombre d'expositions, c'est-à-dire, la distribution des occasions de voir, est elle-même une distribution Bêta Binomiale.

$$(28) \qquad P(S=ki+kj) = \frac{\beta(A+S,\ ni+nj+B-S)}{(ni+nj+1)\beta(A,B)\beta(S+1, ni+nj+1-S)}$$

avec $S = 0,1,2, \ldots, ni + nj$

L'ajustement du modèle, dans l'hypothèse de l'homogénéité des supports, consiste à calculer d'abord la moyenne pondérée des audiences des supports, les poids étant égaux au nombre d'insertions dans chaque support.

$$(29) \qquad \overline{R}_1 = \Sigma_{i=1}^{m}\ A_i ni / \Sigma_{i=1}^{m}\ ni$$

Puis, l'audience de deux parutions du support composite moyen s'obtient en prenant la moyenne pondérée des duplications et des accumulations, les poids étant respectivement égaux au nombre de duplications et au nombre d'accumulations générées par le plan publicitaire considéré. S'il y a m véhicules avec n insertions dans chacun, il y aura $\binom{n}{2} * m$ doubles-expositions dues aux accumulations, et $\binom{m}{2} * n^2$ doubles expositions dues aux duplications, ce qui aboutit à un total de $\binom{nm}{2} = nm(nm-1)/2$ doubles-expositions. D'une manière plus générale, si nous avons un vecteur d'insertions $S = \{n1,n2,\ldots,ni,\ldots,nm\}$, il y aura $\Sigma_{i=1}^{m} \binom{ni}{2}$ accumulations, avec $\binom{ni}{2} = 0$, si $ni < 2$, et $\Sigma_{i=1}^{m}\ \Sigma_{j=i+1}^{m-1}\ ni * nj$

duplications, ce qui fait un total de $\binom{N}{2}$ doubles-expositions avec $N = \Sigma_{i=1}^{m} ni$.

Ainsi, l'audience de deux parutions du support composite moyen est donné par :

$$(30) \qquad \bar{R}_2 = \frac{\left[\Sigma_{i=1}^{m} \binom{ni}{2} R_{2i}\right] + \left[\Sigma_{i=1}^{m} \Sigma_{j=i+1}^{m-1} \; ninjR_{11ij}\right]}{\binom{N}{2}}$$

avec R_{2i} = couverture nette du support i pour deux parutions.

R_{11ij} = couverture nette de la paire de supports i,j.

L'exemple numérique suivant qui utilise le plan publicitaire n° 34 de notre banque de données montre comment appliquer la méthode de Metheringham.

Exemple numérique

Magazines	Audiences	Audiences de 2 parutions	Audiences nettes des Paires de supports		Population
B.H.G.	22 724	30 249	B.H.G.-F.H.	31 116	130 326
F.H.	14 516	30 293	B.H.G.-L.J.H.	32 780	
L.H.J.	16 634	22 598	F.H.-L.H.J.	26 411	
Audiences brutes	53 874	73 140		90 307	

Puisque le plan publicitaire prévoit 2 insertions dans chaque support, il n'y a qu'une accumulation par support et le nombre de duplications est de $2^2 = 4$ pour chaque paire de supports. Ainsi, par application de l'expression (29), nous obtenons l'audience moyenne du support composite :

$$\bar{R}_1 = (53\ 874 \times 2) / 6 = 17\ 958$$

D'après l'expression (30), l'audience de deux parutions du support composite est la suivante :

$$\bar{R}_2 = \left[73\ 140 + (90\ 307 \times 4)\right] / 15 = 28\ 958$$

Nous sommes maintenant en état d'estimer par la méthode des moments les paramètres A et B de la distribution Bêta Binomiale, en utilisant la formule (52) du chapitre 5.

$$\begin{cases} \hat{A} = \dfrac{\bar{R}_1 (\bar{R}_2 - \bar{R}_1)}{POP(2\bar{R}_1 - \bar{R}_2) - \bar{R}_1^2} = 0{,}338 \\[2ex] \hat{B} = \hat{A}\,(POP - \bar{R}_1)/\,\bar{R}_1 = 2{,}115 \end{cases}$$

Enfin, en utilisant l'expression (28), nous pouvons calculer la distribution estimée des occasions de voir.

Modèle de Metheringham : Estimation de la distribution des occasions de voir.

Expositions	Calculs	Estimée	Réelle
0	$\frac{2{,}115*3{,}115*4{,}115*5{,}115*6{,}115*7{,}115}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	79 488	81 087
1	$\frac{6*{,}338*2{,}115*3{,}115*4{,}115*5{,}115*6{,}115}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	22 658	17 315
2	$\frac{15*{,}338*1{,}338*2{,}115*3{,}115*4{,}115*5{,}115}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	12 399	17 465
3	$\frac{20*{,}338*1{,}338*2{,}338*2{,}115*3{,}115*4{,}115}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	7 554	6 504
4	$\frac{15*{,}338*1{,}338*2{,}338*3{,}338*2{,}115*3{,}115}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	4 595	4 995
5	$\frac{6*{,}338*1{,}338*2{,}338*3{,}338*4{,}338*2{,}115}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	2 560	1 708
6	$\frac{0{,}338*1{,}338*2{,}338*3{,}338*4{,}338*5{,}338}{2{,}453*3{,}453*4{,}453*5{,}453*6{,}453*7{,}453}*130\ 326=$	1 077	1 204

Ces estimations sont bien supérieures à celles des modèles à population homogène. La couverture nette réelle n'est surestimée que de 3,2 %. Il semble donc que l'hypothèse de l'homogénéité des supports alliée à une distribution Béta identique pour chacun d'eux n'empêche pas le modèle de produire des esti-

mations acceptables. Néanmoins, les pourcentages d'erreur sont plus élevés pour les niveaux de fréquence non nulle que pour la couverture nette. Cependant, le recours à un hypothétique support moyen conduit à une conséquence extrêmement néfaste. Il s'agit du phénomène de déclin de la couverture nette. Celui-ci apparait lorsque l'adjonction d'un support ou l'accroissement du nombre d'insertions dans un support, provoquent la diminution de la couverture nette du plan, ce qui est, logiquement, impossible. Cela se produit quand le support ou les insertions ajoutées aboutissent à une augmentation du rapport B/(A + B), ce qui se vérifie assez fréquemment.

Ce comportement pathologique n'est pas limité au modèle de Metheringham. Il se produit chaque fois que plusieurs supports sont remplacés par un support moyen hypothétique.

La table suivante résume les performances du modèle de Metheringham pour 57 plans publicitaires.

Résumé des performances du Modèle de Metheringham pour 57 plans publicitaires

Estimation de la couverture nette

Erreur moyenne absolue sur la couverture nette	=	2 483
Pourcentage moyen d'erreur sur la couverture nette	=	5,2 %
Nombre de surestimations de la couverture nette	=	57
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	34
Erreur type de prévision	=	3 410

Estimation de la distribution des occasions de voir

Chi Deux moyen	=	68,77
Nombre de concordances	=	53
Nombre de formes correctes	=	72 (sur 114)

Commentaires

Le modèle Metheringham surestime systématiquement la couverture nette de 5 % en moyenne pour les 57 plans publicitaires. Ce modèle est donc nettement supérieur aux 4 modèles précédents. Plus de la moitié des estimations de la couver-

ture nette sont dans l'intervalle de confiance à 95 % autour des données de l'enquête.

Néanmoins, la surestimation systématique de la couverture nette indique qu'un meilleur modèle peut être construit en abandonnant l'hypothèse de l'homogénéité des supports. Les estimations du modèle sont moins bonnes pour la distribution des occasions de voir. Le nombre de concordances a augmenté considérablement, de 42 pour le meilleur modèle homogène, à 53 pour le modèle de Metheringham. Mais, néanmoins, la distribution Bêta Binomiale ne peut pas reproduire les pointes secondaires observées aux fréquences d'exposition 2 et 4.

Un test plus sévère de la validité de l'ajustement, qui consiste à compter le nombre de fois où la distribution estimée assume la forme correcte aux niveaux d'exposition 2 et 4, indique que le modèle Metheringham n'est correct que dans 72 cas sur 114.

Nos résultats sont similaires à ceux obtenus par Schreiber en 1969 (12) dans son évaluation de la méthode de Metheringham. En résumé, la méthode de Metheringham surestime la couverture nette réelle et les fréquences d'exposition impaires, tandis qu'elle sous-estime les fréquences d'exposition paires. Cela découle directement du fait que la distribution d'exposition estimée est Bêta Binomiale, ce qui implique, soit un mode à la fréquence nulle suivie d'une diminution, soit une forme en U, ou soit un mode à une fréquence intermédiaire conduisant à une forme en cloche. Ainsi, le modèle de Metheringham ne peut pas reproduire les pointes secondaires dues aux lecteurs fidèles. Par exemple, considérons un plan publicitaire comprenant deux supports avec 4 insertions dans chaque. Si la distribution des expositions à chacun des supports est caractérisée par un pic principal pour la fréquence nulle et par un pic secondaire pour 4 expositions, alors, la distribution des expositions au plan publicitaire présentera un pic principal pour la fréquence nulle, un pic secondaire pour 4 expositions, et, éventuellement, un pic tertiaire pour 8 expositions. Le pic principal est dû aux non-lecteurs des deux supports. Le pic secondaire provient des lecteurs fidèles exclusifs de chacun des supports. Le pic tertiaire provient des lecteurs fidèles aux deux supports simultanément.

Plusieurs tentatives ont été faites pour améliorer le modèle de Metheringham. Ces tentatives ont pris deux directions.

1 - Certains chercheurs ont réussi a améliorer les estimations au prix de l'incor-

poration de données supplémentaires autres que les audiences des deux premières parutions et des paires de supports. Néanmoins, cette approche est vouée à l'échec, puisque le but d'un modèle d'exposition est justement d'économiser le temps et les coûts entraînés par de vastes tabulations croisées, de sorte qu'une optimisation, et non pas une simple évaluation des plans publicitaires devienne viable. Cette ligne de recherche comprend le modèle Bêta Matricel, qui sera étudié plus loin, et la formule COMET (13) proposée par Metheringham lui-même. La formule COMET, constitue la base de COSMIC (14), un modèle de planification des supports publicitaires fort populaire au Royaume-Uni depuis son lancement par l'agence Foote, Cone et Belding de Londres en 1970. Bien que ce modèle n'ait fait l'objet d'aucune publication dans les revues spécialisées, on sait qu'outre les audiences des 2 premières parutions et les duplications par paires, la formule COMET nécessite aussi les données suivantes :

- Triplications des supports telles que ABC, ABD,..., etc. C'est-à-dire le nombre d'individus exposés à exactement 3 supports.

- Triplications de 3 parutions telles que AAA, BBB,..., etc. C'est-à-dire le nombre d'individus exposés à exactement 3 parutions du même support.

- Triplications supports-parutions telles que AAB, BBC,..., etc. C'est-à-dire le nombre d'individus exposés à n'importe quel triplet de parutions provenant de deux supports différents.

Cependant, pour 30 supports, il y aura un total de 5 455 observations à emmagasiner dans l'odinateur contre 495 seulement pour la formule de Metheringham originale ! En outre, ces nombres ne proviennent pas tous des données de l'enquête. La plupart d'entre eux doivent d'abord être estimés. En fait, seules les triplications de supports peuvent être obtenues par classification croisées des données de l'enquête. Les audiences cumulées à trois parutions ne sont pas directement mesurées au Royaume-Uni. Ainsi, les triplications de 3 parutions doivent être estimées à partir d'une question d'habitude de lecture. Selon le supplément technique en appendice à la brochure COSMIC, cette estimation est effectuée par la méthode des probabilités de la dernière parution. Ensuite, pour chaque classe de fréquence séparément, les probabilités d'exposition à deux parutions exactement sont estimées par le carré de la probabilité d'exposition à une parution. Ceci revient à admettre l'hypothèse fort peu réaliste d'une accumulation aléatoire à l'intérieur de chacune des classes de fréquence. De même, la probabilité d'exposition à 3 parutions exactement est estimée par le cube de la probabi-

lité d'exposition à une parution, pour chaque classe de fréquence séparément. En ce qui concerne les triplications supports-parutions, elles sont estimées, à partir des duplications par paires et des probabilités d'exposition à deux parutions, sous l'hypothèse d'indépendance. Par exemple :

$$P(AAB) = P(AA) * P(AB)$$

En conclusion, nous estimons que la formule COMET ne représente pas une amélioration réelle de la formule de Metheringham puisqu'elle s'appuie sur des hypothèses irréalistes d'accumulation aléatoire dans chacune des classes de fréquences et d'indépendance entre la duplication des supports et la duplication des parutions. Par ailleurs, les triplications entre supports n'ayant pas été tabulées dans les données I.M.S., dont nous disposons, il ne nous est pas possible de tester empiriquement la formule COMET. Toutefois, en fonction de l'analyse précédente des hypothèses du modèle, nous sommes portés à croire que l'amélioration par rapport à la formule de Metheringham, si elle existe, ne compense certainement pas les coûts engendrés par les tabulations supplémentaires des données de l'enquête et par la mise en mémoire d'un nombre très accru de paramètres.

2 - La seconde voie de recherche s'efforce d'améliorer les estimations tout en ne recourant pas à des données supplémentaires. Ceci peut être obtenu au prix de renoncement à un modèle exact et par l'introduction d'un certain nombre d'approximations Ad Hoc. Les modèles d'Hofmans et de Morgenztern sont un exemple de cette approche, ainsi que le modèle Modal qui, recourant à une hypothèse de linéarité, transforme le problème de l'estimation multivariée en une séquence d'estimation bivariée. Le modèle de casualité constante de Fry procède de façon similaire.

Cependant, il existe encore une troisième possibilité : édifier un modèle exact sans recourir à des données supplémentaires. Nous explorerons cette voie de recherche prometteuse quand nous décrirons notre modèle d'exposition Multinomial Dirichlet.

Après ce panorama du problème et des différentes approches susceptibles de le résoudre, nous examinons maintenant le second modèle à distribution exacte.

Modèle 6 Le modèle Bêta Matriciel

Ce modèle a été présenté en détail par Greene et Stock (15) dans leur brochure "Advertising Reach and Frequency in Magazines". On peut aussi trouver quelques informations supplémentaires dans l'appendice VII des enquêtes d'audience de Simmons (16).

Hypothèses

a) - La population est hétérogène comme dans le modèle 5.

b) - Les supports sont hétérogènes. Chacun d'eux est caractérisé par sa propre distribution Bêta de probabilité d'exposition.

c) - La duplication n'est pas aléatoire. L'accumulation est Bêta Binomiale pour chaque support.

d) - Les probabilités d'exposition sont stationnaires d'une parution à l'autre.

La principale différence entre le modèle Bêta Matriciel et le modèle de Metheringham est que les supports ne sont pas agrégés entre eux pour former un support moyen, mais qu'ils conservent leur propre audience et leur propre taux d'accumulation. Il en est de même pour les duplications par paire qui sont toutes conservées. La principale hypothèse du modèle est que les lignes et les colonnes de la matrice d'exposition suivent une distribution Bêta Binomiale. Chaque ligne et chaque colonne est traitée comme une tranche séparée de population dont les paramètres sont estimés séparément. Par conséquent, quand les paramètres Bêta d'une ligne ou colonne particulière ont été estimés, cette ligne ou cette colonne peut être étendue au nombre désiré d'insertions, comme cela est indiqué au chapître 5, en utilisant la formule (39) de la page 202. Le fonctionnement de la méthode peut être illustré en recourant à un exemple. Considérons la matrice d'exposition Time-Newsweek de 1973.

Matrice d'exposition Time-Newsweek (1973)

(% de la population totale)

		Exposition à Time			Total
		0	1	2	Nswk
Expositions	0	71,4	7,2	6,4	85,0
à	1	4,5	2,8	1,8	9,1
Newsweek	2	2,9	1,3	1,7	5,9
Total Time		78,8	11,3	9,9	100,0

Nous pouvons commencer par estimer, les paramètres Bêta de chaque colonne de la matrice par la méthode des moments. Nous obtenons les résultats suivants :

Paramètres Bêta	Colonnes 0	1	2
$\hat{A}$	0,0573	0,5013	0,2317
$\hat{B}$	0,8292	1,5793	0,6382

En utilisant ces paramètres, nous pouvons estimer, pour chaque colonne, la distribution des occasions de voir pour tout nombre d'insertions désiré. Supposons qu'un plan publicitaire prévoit 5 insertions dans Newsweek et 2 dans Time. La matrice d'exposition correspondante sera la suivante :

5 insertions dans Newsweek et 2 insertions dans Time

(% de la population totale)

		Expositions à Time			
		0	1	2	Total Nswk
Expositions	0	68,2	5,2	5,2	78,6
à	1	4,0	2,4	1,3	7,7
Nexsweek	2	2,2	1,5	0,9	4,6
	3	1,6	1,1	0,8	3,5
	4	1,4	0,7	0,7	2,8
	5	1,4	0,4	1,0	2,8
Total Time		78,8	11,3	9,9	100,0

Maintenant, si un plan publicitaire prévoit 5 insertions dans Newsweek et 3 dans Time, il nous suffit de considérer chaque ligne comme étant une tranche de population séparée ayant sa propre hétérogénéité Bêta. Nous pouvons estimer les paramètres des lignes de la même manière que pour les colonnes.

Paramètres Bêta	LIGNES 0	1	2	3	4	5
$\hat{A}$	0,0583	0,7979	0,8730	0,7596	0,4286	0,1765
$\hat{B}$	0,5291	1,6596	1,5609	1,2097	0,7143	0,2353

En utilisant ces paramètres, les expositions à 3 parutions dans Time sont calculées pour chaque ligne. La matrice d'exposition résultant des 5 insertions dans Newsweek et de 3 insertions dans Time est la suivante :

5 insertions dans Newsweek et 3 insertions dans Time

(% de la population totale)

		Expositions à Time 0	1	2	3	Total Nswk
Expositions à Newsweek	0	66,7	4,6	3,2	4,1	78,6
	1	3,3	2,1	1,5	0,8	7,7
	2	1,8	1,3	0,9	0,6	4,6
	3	1,3	0,9	0,7	0,6	3,5
	4	1,2	0,6	0,5	0,5	2,8
	5	1,3	0,3	0,3	0,9	2,8
Total Time		75,6	9,8	7,1	7,5	100,0

Naturellement, nous aurions pu commencer en développant d'abord Time et ensuite Newsweek. Dans ce cas, nous aurions trouvé une matrice d'exposition légèrement différente parce que l'ajustement Bêta n'est jamais parfait. En outre, si nous développons chaque magazine séparément jusqu'au nombre désiré de parutions, nous obtenons les résultats suivants :

Extension de Time à 3 parutions

Paramètres	$\hat{A}$	$\hat{B}$	Distribution des	0	1	2	3
Bêta	0,1173	0,6345	occasions de voir	75,4	10,1	6,9	7,6

Extension de Newsweek à 5 parutions

Paramètres	$\hat{A}$	$\hat{B}$	Distribution des	0	1	2	3	4	5
Bêta	0,0984	0,8412	occasions de voir	78,5	7,9	4,6	3,4	2,8	2,8

Ainsi, nous observons une légère discordance entre les extensions indépendantes des 2 supports et la ligne et la colonne totale de la matrice d'exposition estimée. Toutefois, les extensions indépendantes doivent être considérées comme étant plus correctes pour deux raisons. D'abord, elles ne présupposent pas que les lignes et les colonnes de la matrice d'exposition soient des distributions Bêta Binomiales. Deuxièmement, chacune d'elles n'entraîne que l'estimation de deux paramètres et, par conséquent, elles sont moins suceptibles d'approximations et d'erreurs d'arrondis que l'estimation de la matrice d'exposition qui, dans le cas traité, exige l'estimation de 6 paramètres pour les colonnes, puis de 10 paramètres pour les lignes.

Les discordances qui sont fort réduites dans notre exemple à 2 supports tendraient à augmenter pour des plans publicitaires plus étendus, puisque les diverses manières de réaliser l'extension croissent exponentiellement. Pour m supports, l'extension peut être faite de 2^m manières différentes. Greene et Stock surmontent cette difficulté en prenant un échantillon selon un carré latin systématique, de toutes les extensions possibles et en faisant la moyenne. Evidemment, ce procédé implique des calculs considérables et exige beaucoup de temps. En outre, ce recours à une moyenne ne garantit pas que les marges de la matrice d'exposition seront égales aux extensions indépendantes de chaque support. Ainsi, Greene et Stock utilisent l'algorithme de Deming (17) pour ajuster les colonnes et les lignes de la matrice d'exposition jusqu'à ce qu'elles reproduisent parfaitement les extensions indépendantes de chaque support.

En appliquant cet algorithme à la matrice d'exposition estimée de la page précédente, nous obtenons la matrice finale d'exposition à 5 parutions de Newsweek et à 3 parutions de Time.

Matrice d'Exposition Ajustée

5 Insertions dans Newsweek et 3 Insertions dans Time
(en % de la population adulte)

		Nombre d'expositions à Time				
		0	1	2	3	Total pour Newsweek
Nombre d'expositions à Newsweek	0	66,46	4,72	3,10	4,22	78,5
	1	3,38	2,21	1,50	0,81	7,9
	2	1,80	1,34	0,87	0,59	4,6
	3	1,26	0,90	0,66	0,58	3,4
	4	1,20	0,62	0,48	0,50	2,8
	5	1,30	0,31	0,29	0,90	2,8
Total pour Time		75,5	10,1	6,9	7,6	100,0

En additionnant les éléments de la matrice le long de la diagonale montante, nous obtenons la distribution des occasions de voir générée par 5 insertions dans le magazine Newsweek et 3 insertions dans le magazine Time.

Distribution des Occasions de Voir

Expositions	0	1	2	3	4	5	6	7	8	Couverture Nette
% de la population	66,46	8,10	7,11	8,32	3,78	3,17	1,37	0,79	0,90	33,54

Nous observons que la méthode Bêta Matricielle conserve les pics intermédiaires de la distribution des occasions de voir dus aux lecteurs fidèles. Le pic, au niveau de 3 expositions, provient des lecteurs fidèles de Time. Le faible pic, au niveau de 8 expositions, est dû aux lecteurs qui sont fidèles aux deux magazines, simultanément.

A titre de comparaison, nous allons montrer que la méthode de Metheringham, appliquée aux mêmes données, fait disparaître les pics intermédiaires. En utilisant la matrice d'exposition Time-Newsweek de la page 468 nous pouvons calculer les 5 audiences suivantes en pourcentage de la population

totale.

Magazine	1ère parution	2ème parution	Audience nette de la paire Time-Newsweek
TIME	15,55	21,2	22,05
NEWSWEEK	10,45	15,0	

Il convient de noter que ces 5 audiences sont les seules prises en compte par la méthode de Metheringham. Elles ne proviennent que de 4 éléments parmi les 9 de la matrice d'exposition. Le calcul des paramètres A et B, correspondant à 5 insertions dans Newsweek et à 3 insertions dans Time, s'effectue selon la formule (30) de la page 461 .

$$\bar{R}_1 = (15{,}55 * 3 + 10{,}45 * 5) / 8 = 12{,}3625$$

$$\bar{R}_2 = [(21{,}2 * 3 + 15{,}0 * 10) + (22{,}05 * 15)] / 28 = 19{,}4411$$

$$\hat{A} = 0{,}233 \qquad \hat{B} = 1{,}6518$$

A partir de ces deux paramètres, nous obtenons la distribution estimée des occasions de voir par la formule (28) de la page 460 .

Distribution des Occasions de voir par la méthode de Metheringham

Expositions	0	1	2	3	4	5	6	7	8	Couverture nette
% de la population	63,24	13,63	7,68	5,16	3,69	2,69	1,92	1,29	0,7	36,76

Nous observons que la méthode de Metheringham conduit à une couverture nette supérieure à celle de la méthode Bêta Matricielle et que, de plus, elle fait disparaître les pics intermédiaires au niveau de 3 et de 8 expositions. Plus grave encore, la méthode de Metheringham se montre incapable de reproduire la matrice des données originales. Avec deux insertions dans chaque support, la couverture nette et la distribution des occasions de voir sont estimées comme suit :

$$\bar{R}_1 = (15{,}55 * 2 + 10{,}45 * 2) / 4 = 13{,}0$$

$$\bar{R}_2 = (21{,}2 + 15{,}0 + 22{,}05 * 4)\ 16 = 20{,}7333$$

$$\hat{A} = 0{,}2811 \qquad \hat{B} = 1{,}8811$$

Distribution des Occasions de voir : 2 Insertions dans Time et Newsweek

Expositions	0	1	2	3	4	Couverture nette
Observées	71,40	11,70	12,10	3,10	1,70	28,6
Metheringham	69,89	16,10	7,97	4,21	1,83	30,11

Ainsi, la méthode de Metheringham ne reproduit pas la matrice d'exposition à 2 parutions et la distribution des occasions de voir, car elle surestime la couverture nette et fait disparaître le pic des lecteurs fidèles au niveau de 2 expositions. Hélas, ceci ne constitue pas encore la plus grave critique que l'on puisse adresser au modèle de Metheringham. Non seulement le modèle échoue sur le critère de reproduction des données originales, mais encore, il échoue également sur le critère d'unicité de la solution. En fait, plusieurs matrices d'exposition correspondant à des taux d'accumulation et de duplication différents conduisent à des résultats identiques, comme nous allons le montrer maintenant. Nous adopterons la notation suivante :

R_{1i} = Audience d'une parution du support i.

R_{1j} = Audience d'une parution de support j.

R_{2i} = Audience cumulée de deux parutions de support i.

R_{2j} = Audience cumulée de deux parutions du support j.

R_{11ij} = Couverture nette de la paire de supports i et j.

Les équations de la méthode de Metheringham qui conduisent à l'estimation des paramètres A et B du support moyen, dans le cas de 2 insertions dans chaque support, sont les suivantes :

$$\begin{cases} \overline{R}_1 = (R_{1i} + R_{1j}) / 2 \\ \overline{R}_2 = (R_{2i} + R_{2j} + 4R_{11ij}) / 6 \end{cases}$$

Il existe une infinité de valeurs R_{1i}, R_{1j}, R_{2i}, R_{2j}, R_{11ij}, conduisant aux mêmes valeurs de $\overline{R}_1$ et $\overline{R}_2$. En d'autres termes, le modèle de Metheringham ne vérifie pas la propriété d'unicité de la solution puisque des données différentes

peuvent conduire à une même estimation dès lors qu'elles vérifient le système d'équation précédent. A titre d'exemple, les deux matrices d'exposition suivantes produisent la même couverture nette, 30,11 %, et la même distribution des occasions de voir que la matrice Time-Newsweek de la page 468 . Cependant, dans la première matrice, la duplication est plus élevée et l'accumulation plus faible, tandis que dans la seconde, la duplication est plus faible et l'accumulation plus élevée.

Non Univocité du Modèle de Metheringham

2 Matrices d'Exposition différentes produisant la même couverture nette

Matrice d'Exposition N° 1

	0	1	2	Total
0	73,6	3,4	11,8	88,8
1	1,4	0,0	0,0	1,4
2	7,8	0,0	2,0	9,8
Total	82,8	3,4	13,8	100,0

Matrice d'Exposition N° 2

	0	1	2	Total
0	72,6	7,0	1,2	80,8
1	1,8	12,0	3,6	17,4
2	0,4	0,4	1,0	1,8
Total	74,8	19,4	5,8	100,0

La matrice 1, la matrice 2 et la matrice Time-Newsweek produisent toutes les trois une couverture nette estimée de 30,11 % alors que les couvertures nettes réelles sont respectivement de 26,4 %, 27,4 % et 28,6 % !

La méthode Bêta Matricielle est exempte de toutes les faiblesses de la méthode de Metheringham. De plus, lorsque l'on compare ses estimations avec les résultats observés par tabulation croisée des données, les différences sont toujours extrêmement faibles et les pics secondaires de la distribution des occasions de voir sont toujours reproduits avec fidélité. Ceci ne peut pas être montré à l'aide de notre banque de données puisqu'elle est limitée à deux parutions dans chaque support. Toutefois, certaines enquêtes, notamment celle menée en 1964 par Simmons, portent sur 4 parutions et permettent donc de tester la méthode Bêta Matricielle. Les résultats de ce test sont présentés dans les deux tables suivantes.

Test du Modèle Bêta Matriciel
par Tabulation croisée des audiences de 4 parutions
mesurées par Simmons en 1964

(Population en milliers d'adultes masculins)

	LOOK-POST		LOOK-LIFE		LIFE-POST		LIFE-LOOK-POST	
Expositions	OBS	EST	OBS	EST	OBS	EST	OBS	EST
0	28 470	28 436	25 268	25 456	26 702	26 851	21 131	21 255
1	4 531	4 973	5 272	5 014	5 605	5 394	4 688	4 698
2	5 187	4 133	4 646	4 372	4 810	4 507	4 834	4 195
3	3 741	4 538	4 344	4 846	4 310	4 922	4 024	4 693
4	8 825	9 864	8 493	8 240	9 522	9 296	8 337	8 105
5	1 908	1 890	2 348	2 639	1 769	1 792	2 530	2 791
6	1 337	1 275	1 899	1 826	1 249	1 193	2 431	2 258
7	1 185	1 124	1 870	1 618	1 069	1 054	2 229	2 171
8	2 343	2 393	3 387	3 523	2 491	2 518	3 885	3 835
9							781	986
10							734	668
11							685	577
12							1 238	1 304
Couverture nette	29 057	29 091	32 259	32 080	30 825	30 675	36 396	36 281

Test du Modèle Bêta Matriciel * (suite)

par Tabulation croisée de 6 parutions des magazines LIFE et LOOK mesurés par Politz en 1958

(Population en milliers de foyers)

Expositions	Observations	Estimations	Différences	% d'Erreur
0	37 400	39 319	+1 919	5,1
1	16 299	13 164	-3 135	-19,2
2	11 498	10 707	- 791	- 7,0
3	10 234	9 955	- 279	- 2,7
4	9 224	9 833	+ 609	+ 6,6
5	8 465	10 265	+1 800	+21,2
6	9 982	12 559	+2 577	+25,8
7	5 812	5 703	- 109	- 1,8
8	5 686	4 017	-1 669	-29,3
9	3 664	3 090	- 574	-15,6
10	3 032	2 529	- 503	-16,6
11	2 780	2 277	- 503	-18,0
12	2 274	2 934	+ 660	+29,0
Couverture nette	88 950	87 031	-1 919	- 2,1

* Source : Schreiber R.J. (12)

Ils sont, en vérité, excellents tant du point de vue de la couverture nette qui est exempte de tout biais systématique, que du point de vue de la distribution des occasions de voir où les pics sont bien reproduits, quoique l'on puisse déceler une légère tendance à leur accentuation.

Il semble donc que nous soyons enfin en possession d'une méthode d'estimation parfaite de la couverture nette et de la distribution des occasions de voir. En un sens oui et en un sens non. Il est vrai que la méthode reproduit extrêmement bien les données et qu'elle est exempte du phénomène de déclin de la couverture nette puisqu'elle ne fait pas appel à un hypothétique support moyen. Toutefois, la méthode requiert la connaissance de la matrice d'exposition à deux parutions. Or, cette matrice ne peut être obtenue que par la tabulation croisée des audiences de chacun des supports inclus dans le plan publicitaire. Par exemple, dans le cas de 3 supports, cela nécessite une tabulation 3 x 3 x 3 pour obtenir les 27 éléments de la matrice d'exposition tri-dimensionnelle. Le coût de ces tabulations croisées devient très rapidement prohibitif dès lors qu'il faut considérer plus de six ou sept supports. D'une façon générale, si le nombre de supports inclus dans le plan publicitaire est de m, il y aura 3^m éléments à tabuler pour former la matrice d'exposition. Inévitablement, un certain nombre de ces éléments ne comprendra aucun des individus de l'échantillon ou bien un nombre si faible qu'il sera nécessairement sujet à caution. De plus, si le plan publicitaire prévoit un grand nombre d'insertions dans un grand nombre de supports, le nombre de paramètres Bêta à estimer et le nombre d'extrapolations à effectuer deviendra tout à fait gigantesque. Par exemple, un plan publicitaire très modeste ne prévoyant que 4 insertions dans 4 supports requiert l'estimation de 580 paramètres Bêta et autant d'extrapolations. La matrice d'exposition aura 2 401 éléments et elle devra être ajustée aux 4 extensions indépendantes de chacun des supports, ce qui nécessitera de lourds calculs. En conclusion, d'un point de vue pratique, la méthode est limitée aux petits plans publicitaires ne dépassant pas six ou sept supports avec un nombre modeste d'insertions dans chacun d'eux.

A cause de cette impasse, les chercheurs ont tourné leurs espoirs soit vers les modèles de simulation, soit vers des modèles faisant appel à des approximations. L'objectif à atteindre est clairement défini. Il s'agit d'améliorer les estimations du modèle de Metheringham tout en ne faisant pas appel à des données supplémentaires. Les deux approches, simulation et approximation, ont conduit à la réalisation de cet objectif, mais très peu d'informations détaillées ont été publiées à ce jour sur la manière de l'atteindre. Ceci est

compréhensible étant donné la valeur commerciale de ces modèles. Avant de décrire les principes et d'analyser les performances de ces modèles, nous allons présenter un nouveau modèle à distribution exacte qui ne requiert d'autres données que celles utilisées par le modèle de Metheringham. Nous expliquerons pourquoi ce modèle a éludé les efforts des chercheurs précédents et nous montrerons qu'il remédie aux trois défauts majeurs du modèle de Metheringham; le déclin de la couverture nette, la non unicité de la solution et la reproduction des données originales.

Modèle 7 Le Modèle Multinomial Composé (18)

Ce nouveau modèle est une extension du modèle de Metheringham obtenue en levant l'hypothèse d'homogénéité des supports. En effet, c'est cette hypothèse irréaliste qui est à l'origine des défauts du modèle de Metheringham.

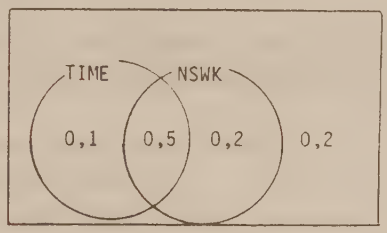
Hypothèses :

a) - La population est hétérogène.
b) - Les supports sont hétérogènes. Chaque support est caractérisé par sa propre distribution des probabilités individuelles d'exposition qui est du type Bêta. Toute combinaison de supports est donc caractérisée par une distribution Bêta multivariée.
c) - La duplication est non aléatoire. L'accumulation est Bêta Binomiale pour chacun des supports.
d) - Les probabilités d'exposition sont stationnaires d'une parution à une autre.

Ces hypothèses sont similaires à celles du modèle 6; toutefois, les données requises par le modèle sont identiques à celles du modèle 5. L'idée d'utiliser la distribution Bêta Multivariée, aussi connue sous le nom de distribution de Dirichelet, pour représenter l'hétérogénéité des supports et des individus n'est pas entièrement nouvelle. Grenne et Stock (14) évoquent cette possibilité mais ils la rejettent sur la base du raisonnement suivant : "..., il faudrait admettre que chaque individu possède une probabilité d'exposition à chaque support plus une probabilité résiduelle de non exposition à l'ensemble des supports. Chacune de ces probabilités doit être non-négative et leur somme doit être égale à l'unité. Donc, la distribution de Dirichelet ne peut pas s'appliquer aux individus tels que la somme de leurs probabilités d'exposition à un ensemble de supports donnés soit supérieure à l'unité". Greene et Stock

se réfèrent à la possibilité qu'un individu possède, par exemple, une probabilité d'exposition de 0,6 à Time et de 0,7 à Newsweek. Cette possibilité est réelle, mais en fait, elle ne fait pas problème car il suffit de redéfinir les évènements à probabiliser de façon à ce que la somme de leurs probabilités soit toujours égale à l'unité, quel que soit l'individu et quel que soit le nombre de supports. Plutôt que de raisonner comme le font Greene et Stock, en termes de probabilité d'exposition non exclusive à un support, il suffit de raisonner en termes de probabilité d'exposition exclusive à ce support. L'exemple suivant clarifie cette distinction dans le cas de deux supports et il est aisément généralisable à plusieurs supports. Supposons qu'un individu soit caractérisé par une probabilité d'exposition à Time de 0,6, une probabilité d'exposition à Newsweek de 0,7 et une probabilité de non exposition à ces deux supports de 0,2. La somme de ces probabilités n'est pas égale à l'unité car les trois évènements auxquels elles se réfèrent ne sont pas mutuellement exclusifs, comme le montre le schéma suivant et la matrice d'exposition correspondante.

		Time: Non Exp.	Time: Exp.	Expo. à Newsweek
Newsweek	Non Exp.	0,2	0,1	0,3
	Exposit.	0,2	0,5	0,7
Exp. à Time		0,4	0,6	1,0



Modèle Binomial Bivarié	
Evènements non-exclusifs	Probabilités
Exposition à Time	0,6
Exposition à Newsweek	0,7
Non exposition aux deux supports	0,2
Total	1,5

Modèle Multinomial	
Evènements exclusifs	Probabilités
Exposition à Time seul	$0,1 = p_1$
Exposition à Newsweek seul	$0,2 = p_2$
Exposition aux deux	$0,5 = p_3$
Non exposition aux deux	$0,2 = p_0$
Total	1,0

Il est donc clair que l'objection de Greene et Stock ne s'applique pas lorsque l'on considère les expositions exclusives. Nous allons présenter le modèle Multinomial composé en raisonnant sur le cas de deux supports. Si nous supposions un instant que les probabilités p_0, p_1, p_2, p_3 soient identiques pour

tous les individus-cible, la probabilité d'obtenir un vecteur d'exposition avec k0 non expositions, k1 expositions à Time seul, k2 expositions à Newsweek seul et k3 expositions aux deux magazines, serait donnée par la distribution multinomiale suivante :

$$(31) \qquad P(k0,k1,k2,k3) = \frac{n!}{k0!k1!k2!k3!} \; p_0^{k0} \; p_1^{k1} \; p_2^{k2} \; p_3^{k3}$$

$$\text{avec } \Sigma_{i=0}^{3} \; ki = n, \qquad ki = 0, 1, 2, \ldots, n$$

Mais, en vertu de l'hypothèse b, les probabilités p_i sont elles-même des variables aléatoires suivant une distribution Bêta Multivariée (19), définie par :

$$(32) \quad P(p_0,p_1,p_2,p_3) = \frac{\Gamma(A0+A1+A2+A3)}{\Gamma(A0)\Gamma(A1)\Gamma(A2)\Gamma(A3)} \; p_0^{A0-1} p_1^{A1-1} p_2^{A2-1} p_3^{A3-1}$$

$$\text{avec } 0 \leq p_i \leq 1 \; ; \; \Sigma_{i=0}^{3} \; p_i = 1 \; ; \; A0, A1, A2, A3 > 0$$

En composant la distribution (31) avec la distribution (32), nous obtenons la distribution Bêta Multinomiale ou Multinomiale Composée (20).

Soit K, le vecteur d'exposition : $K = (k0, k1, k2, k3)$
p, le vecteur des probabilités : $P = (p_0, p_1, p_2, p_3)$
et A, le vecteur des paramètres : $A = (A0, A1, A2, A3)$

nous obtenons :

$$(33) \quad P(K/A,n) = \frac{n!}{k0!k1!k2!k3} * \frac{\Gamma(\Sigma_{i=0}^{3} Ai) * \pi_{i=0}^{3} \Gamma(ki+Ai)}{[\pi_{i=0}^{3} \Gamma(Ai)] \; [\Gamma(n + \Sigma_{i=0}^{3} Ai)]}$$

$$\text{avec } \Sigma_{i=0}^{3} \; ki = n \; ; \qquad ki = 0, 1, 2, \ldots. n \; ; \; Ai > 0 \; \forall \; i.$$

L'expression (33) peut être simplifiée comme suit pour faciliter les calculs

$$(34) \; P(K/A,n) = \frac{n!}{k0!k1!k2!k3!} * \frac{[A0(A0+1)..(A0+k0-1)] \ldots. [A3(A3+1)..(A3+k3-1)]}{(A0+A1+A2+A3)(A0+A1+A2+A3+1)\ldots(A0+A1+A2+A3+n-1)}$$

Le modèle Multinomial composé peut aussi être généré, de façon moins abstraite, par analogie entre l'exposition à un support et le tirage de boules à partir

d'une urne. L'ensemble de m supports peut être assimilé à une urne contenant des boules de m couleurs différentes, avec A_i boules de la ième couleur. Les boules de couleurs i sont assimilables aux individus constituant l'audience exclusive du support i. Si une boule de couleur i est tirée, puis remise dans l'urne avec c boules de la même couleur, ceci définit le schéma de tirage de Polya-Eggenberger. Quand c = 1, la distribution du nombre de boules tirées dans chaque couleur est donnée par la distribution Multinomiale composée, comme nous allons le montrer. La probabilité d'obtenir le vecteur $K = (k0, k1, \ldots, ki, \ldots km)$ où ki représente le nombre de boules de couleur i obtenues en n tirages, est donnée par la distribution de Polya multivariée :

$$(35) \quad p(K) = \frac{n!}{\pi_{i=0}^{m} ki!} * \frac{\pi_{i=0}^{m} A_i^{(ki,c)}}{\left[\Sigma_{i=0}^{m} A_i\right]^{(n,c)}}$$

$$\text{avec } A_i^{(ki,c)} = A_i\ (A_i + 1) \ldots \quad [A_i + (ki-1)c]$$

dans le cas où $c = 1$, l'expression (35) se simplifie comme suit :

$$(36) \quad P(K) = \frac{n!}{\pi_{i=0}^{m} ki!} * \frac{[A0(A0+1)..(10+k0-1)] \ldots [Am(Am+1)..(Am+k-1)]}{(\Sigma_{i=0}^{m} A_i)\ (1 + \Sigma_{i=0}^{m} A_i)\ldots(n-1 + \Sigma_{i=0}^{m} A_i)}$$

$$\text{avec } \Sigma_{i=0}^{m} ki = n\ ;\ ki = 0, 1, 2, \ldots, n)\ \ A_i > 0\ \forall\ i$$

L'expression (36) n'est autre que l'expression (34) étendue à m supports et la démonstration est ainsi achevée.

La distribution multinomiale composée possède plusieurs propriétés intéressantes. Tout d'abord, lorsqu'un seul support est considéré, l'expression (36) se réduit à la distribution Bêta Binomiale. Donc, la distribution Multinomiale composée peut être considérée comme étant la généralisation à m supports de la loi Bêta Binomiale. Deuxièmement, la distribution jointe de tout sous-ensemble de supports est aussi une distribution multinomiale composée et, par conséquent, les distributions conditionnelles et les distributions marginales sont également des distributions Bêta Binomiales. Cette propriété, qui constituait l'une des hypothèses du modèle Bêta Matriciel et que nous avions vérifiée de façon empirique, trouve donc ici une justification théorique. Si nous appelons K-i le vecteur d'exposition à tous les supports à l'exception du support i, la distribution conditionnelle des expositions au support i est donnée par :

$$(37)\quad P(ki/K-i) = \frac{n!}{ki!\ \Sigma K!} * \frac{[A_i(A_i+1)\ \ldots(A_i+ki-1)]\ [B(B+1)\ldots(B+\Sigma K-1)]}{(A_i+B)(A_i+B+1)\ldots(A_i+B+n-1)}$$

avec $\Sigma K = \Sigma_{j=0}^{m} kj - ki$, et $B = \Sigma_{j=0}^{m} A_j - A_i$.

Par exemple, dans le cas de deux supports, la distribution des probabilités d'exposition au support 1, pour k0 non expositions aux deux supports, k2 expositions exclusives au support 2 et k3 expositions aux deux supports, est Bêta Binomiale de paramètres A_1 et B :

$$(38)\quad P(k1/k0,k2,k3) = \binom{n}{k1} \frac{[A1(A1+1)\ldots(A1+k1-1)]\ [B(B+1)\ldots(B+n-k1-1)]}{(A1+B)\ (A1+B+1)\ldots(A1+B+N-1)}$$

avec $B = \Sigma_{j=0}^{m} Aj - A1$.

La plus intéressante propriété du modèle est que toute somme de variables multinomiales composées est encore une variable multinomiale composée. Ceci provient du fait qu'une somme de variables multinomiales est une variable multinomiale, de même qu'une somme de variables de Dirichlet est encore une variable de Dirichlet (21), (22). Par exemple, dans le cas de deux supports, la somme du nombre d'expositions exclusives au support 1 et du nombre d'expositions exclusives aux deux supports, c'est-à-dire en fait le nombre d'expositions non exclusives au support 1, est une variable multinomiale composée de paramètres α et β, avec $\alpha = A1 + A3$ et $\beta = A0 + A2$. Donc, la distribution marginale des expositions au support 1 est Bêta Binomiale de paramètres α et β. Il convient donc de noter que, non seulement le modèle Multinomial composé se réduit au modèle Bêta Binomial quand il n'existe qu'un seul support, mais encore, les distributions marginales de chaque support sont Bêta Binomiale lorsque plusieurs supports sont considérés. Dans le cas de 2 supports, la distribution marginale d'exposition au support 1 est donnée par :

$$(39)\quad P(K1) = \binom{n}{K1} \frac{[\alpha\ (\alpha+1)\ \ldots\ (\alpha+K1-1)]\ [\beta\ (\beta+1)\ldots(\beta+n-k1-1)]}{(\alpha+\beta)(\alpha+\beta+1)\ldots\ (\alpha+\beta+n-1)}$$

avec K1 = 0,1, ..., n = Nombre d'expositions non exclusives au support 1 et avec $\alpha = A1 + A3$; $\beta = A0 + A2$.

Pour obtenir la distribution marginale des expositions au support 2, il suffit de prendre $\alpha = A2 + A3$ et $\beta = A0 + A1$.

Finalement, la distribution des occasions de voir, générée par le plan publicitaire est obtenue en sommant l'équation (34) pour toutes les valeurs non négatives de ki telles que k1 + k2 + 2k3 = S, le nombre d'expositions au plan publicitaire.

(40) $$P(S) = \sum_{S} P(K/A,n) \quad \text{avec} \quad S = k1 + k2 + 2k3, \quad S = 0, 1, \ldots, 2n$$

et $K = \{k0,k1,k2,k3\}$ $\quad ki = 0,1, \ldots, n$

$A = \{A0,A1,A2,A3\}$ $\quad A_i > 0 \ \forall \ i.$

Cette formule peut être étendue aisément au cas de m supports en utilisant l'expression (36).

Envisageons maintenant le problème de l'estimation des paramètres du modèle Multinomial composé. Mosimann(20) a établi les moments de la distribution Multinomiale composée. On les trouve aussi dans l'ouvrage de Johnson et Kotz (21) et nous les avons reproduits ci-dessous.

(41) $$E(ki = nAi/s \qquad \text{avec } s = \Sigma_{i=0}^{m} Ai$$

(42) $$E(kikj) = n(n-1)\ AiAj\ /\ [s(1 + s)]$$

(43) $$Var(ki) = nAi(s-Ai)\ (n + s)\ \ /\left[s^2(1 + s)\right]$$

(44) $$Cov(kikj) = -nAiAj(n + s)\ /\ \left[s^2(1 + s)\right]$$

(45) $$\rho_{ij} = \left[-\ AiAj\ /\ [(s-Ai)\ (s-Aj)]\right]^{1/2}$$

On peut noter que les variances et les covariances sont similaires à celles du modèle Multinomial, exception faite pour le facteur: $C = (n + s) / (1 + s)$. On notera aussi que la régression multiple de ki sur k-i est linéaire :

(46) $$E(ki/k\text{-}i) = (n - \Sigma_{j=0}^{m} kj - ki)ki/\ (\Sigma_{j=0}^{m} Aj - Ai)$$

Ce fait rend crédible la méthode d'approximation utilisée par le modèle Modal, que nous étudierons par la suite.

En égalisant les moments de la distribution Multinomiale composée avec leur valeur observée dans l'échantillon, on obtient l'estimation, par la méthode des moments, du vecteur des paramètres.

$$(47) \quad \hat{A}i = \bar{K}i\,(n - \hat{c}) \,/\, n(\hat{c} - 1) \quad , \quad i = 1, 2, \ldots, m$$

Ce système d'équations peut être résolu si l'on dispose au préalable d'une estimation de c. Mosimann propose d'estimer la racine k-1 ième de c par le rapport du déterminant de la matrice de variance-covariance Multinomiale composé au déterminant de la matrice de variance-covariance Multinomiale simple.

$$\hat{c} = \left[\frac{|\,W_{Mc}\,|}{|\,W_M\,|}\right]^{1/k-1}$$

Malheureusement, cette méthode n'est pas applicable aux données dont nous disposons. Nous avons donc dû mettre au point une méthode d'estimation du vecteur des paramètres qui ne requiert que la connaissance des audiences simples et cumulées à 2 parutions ainsi que l'audience nette de chaque paire de supports. En se reportant à la définition de c, il est évident que, puisque la valeur n du nombre d'expositions est connue, c ne dépend que de s, la somme des paramètres à estimer. Notre solution consiste donc à estimer d'abord s, puis à en déduire la valeur des paramètres A_i. L'expression (39) montre que la somme des paramètres α et β de la distribution marginale de chacun des supports doit être constante et égale à s si le modèle multinomial composé s'applique. En fait, on peut considérer qu'il s'agit là d'un test d'applicabilité du modèle. Dans la pratique, si les supports diffèrent quant à la somme de leurs paramètres Bêta, il faudra recourir à une méthode d'estimation de s plus complexe. En utilisant les expressions (41) et (34), nous obtenons dans le cas de 2 supports les 8 équations suivantes qui permettent d'estimer s en fonction des audiences cumulées à 2 parutions de chaque support et de l'audience nette de la paire de supports :

$$\begin{cases} A0/s = p0 \\ A1/s = P1 \\ A2/s = p2 \\ A3/s = p3 \end{cases} \qquad \begin{cases} (A0+A2)(A0+A1+1) \,/\, s\,(s+1) = 1 - R_{21} \\ (A0+A1)(A0+A1+1) \,/\, s\,(s+1) = 1 - R_{22} \\ (A1+A3)(A1+A3+1) \,/\, s\,(s+1) = 2R_{11} - R_{21} \\ (A2+A3)(A2+A3+1) \,/\, s\,(s+1) = 2R_{12} - R_{22} \end{cases}$$

avec : $s = A0 + A1 + A2 + A3.$

R_{11} = Audience d'une parution du support 1.

R_{21} = Audience de deux parutions du support 1.

R_{12} = Audience d'une parution du support 2.

R_{22} = Audience de deux parutions du support 2.

En résolvant ces équations, on obtient :

$$\frac{2\,s + 4}{s + 1} = \frac{1 - R_{21}}{1 - R_{11}} + \frac{1 - R_{22}}{1 - R_{12}} + \frac{2\,R_{11} - R_{21}}{R_{11}} + \frac{2\,R_{12} - R_{22}}{R_{22}}$$

En appelant z le membre de droite de cette équation, l'estimation de s est donnée par :

$$\hat{s} = (4 - z) / (z - 2)$$

Une fois s estimé, on obtient chacun des paramètres par :

$$\hat{A}_i = p_i * \hat{s} \qquad i = 0, 1, 2, 3.$$

En résumé, nous avons présenté un nouveau modèle d'exposition complet, basé sur une distribution exacte et ne requérant pas d'autres données que celles nécessaires au modèle de Metheringham. Nous allons maintenant montrer que les performances empiriques de ce nouveau modèle sont supérieures à celles du modèle de Metheringham. L'exemple numérique suivant décrit en détail l'application du modèle Multinomial composé à la matrice d'exposition Time-Newsweek utilisée précédemment pour le modèle de Metheringham.

Exemple numérique

Les données nécessaires sont identiques à celles utilisées par le modèle de Metheringham.

Magazines	1 parution	2 parutions	Audience nette de la paire de supports
Time	15,55	21,2	
Newsweek	10,45	15,0	TIME-NSWK = 22,05

A partir de ces données, nous calculons les 4 probabilités d'exposition exclusive à la paire de supports.

Non exposition aux 2 supports	:	p_0	$= 1 - 0{,}2205$	$= 0{,}7795$
Exposition exclusive à Time	:	p_1	$= 0{,}2205 - 0{,}1047$	$= 0{,}1158$
Exposition exclusive à Newsweek	:	p_2	$= 0{,}2205 - 0{,}156$	$= 0{,}0645$
Exposition aux 2 supports	:	p_3	$= 0{,}156 - 0{,}1158$	$= 0{,}0402$

Nous estimons ensuite les paramètres Bêta de chacun des supports.

<u>Time</u> :

$$A_T = \frac{0{,}156\ (0{,}2125 - 0{,}156)}{2 * (0{,}156)-(0{,}156)^2 - 0{,}2125} = 0{,}1173$$

$$B_T = \frac{(0{,}1173)(1 - 0{,}156)}{0{,}156} = 0{,}6344$$

<u>Newsweek</u> :

$$A_N = \frac{(0{,}1047\ (0{,}1501 - 0{,}1047)}{2 * (0{,}1047)-(0{,}1047)^2-0{,}1501} = 0{,}0983$$

$$B_N = \frac{(0{,}9833)(1 - 0{,}1047)}{0{,}1047} = 0{,}8409$$

Nous vérifions que la somme des paramètres de Time, $A_T + B_T = 0{,}7517$, est suffisamment proche de la somme des paramètres de Newsweek, $A_N + B_N = 0{,}9392$, pour que le modèle Multinomial composé puisse s'appliquer.

L'estimation de s nous donne, $\hat{s} = 0{,}827$, et, par conséquent, le vecteur des paramètres est estimé comme suit :

$$\begin{cases} \hat{A}0 = p_0 * \hat{s} = 0{,}6446 \\ \hat{A}1 = p_1 * \hat{s} = 0{,}0958 \\ \hat{A}2 = p_2 * \hat{s} = 0{,}533 \\ \hat{A}3 = p_3 * \hat{s} = 0{,}0333 \end{cases}$$

Nous pouvons maintenant calculer, par l'expression (36), la distribution jointe des expositions, la matrice d'exposition et la distribution des occasions de voir, comme le montre la table suivante :

Modèle Multinomial Composé
2 Insertions dans Time et Newsweek

Evènements				Probabilités	Matrice d'Expositions		Fréquence d'exposition	
k0	k1	k2	k3		TIME	NSWK	Fréquence	Probabilité
2	0	0	0	0,7016	0	0	0	0,7016
1	1	0	0	0,0817	1	0	1	0,1272
1	0	1	0	0,0455	0	1		
0	2	0	0	0,0695	2	0	2	0,1418
0	0	2	0	0,0372	0	2		
0	1	1	0	0,0067 } 0,0351	1	1		
0	0	0	1	0,0284				
0	1	0	1	0,0042	2	1	3	0,0066
0	0	1	1	0,0024	1	2		
0	0	0	2	0,0228	2	2	4	0,0228

La qualité des estimations du modèle peut être appréciée en comparant la matrice d'exposition réelle avec la matrice d'exposition estimée. La table suivante présente cette comparaison et souligne l'amélioration obtenue par rapport au modèle de Metheringham. Le test d'ajustement du Chi Deux est 8 fois meilleur pour le modèle multinomial composé que pour le modèle de Metheringham. De même, la distribution des fréquences est nettement mieux reproduite par le modèle Multinomial Composé. Le pic secondaire observé pour 2 expositions est correctement reproduit alors qu'il est totalement effacé par le modèle de Metheringham. Enfin, l'estimation de la couverture nette est sensiblement améliorée.

Comparaison des Performances du Modèle Multinomial Composé et du Modèle de Metheringham

(Pour chaque case du tableau, la 1ère ligne donne la fréquence observée, la 2ème ligne, l'estimation multinomiale composée et la 3ème ligne l'estimation du modèle de Metheringham).

		EXPOSITION à TIME			Total NSWK
		0	1	2	
EXPOSITION à NEWSWEEK	0	71,42 70,16 69,89	7,18 8,17 8,05	6,39 6,95 1,33	84,98 85,28 79,27
	1	4,48 4,55 8,05	2,79 3,51 5,31	1,81 0,42 2,105	9,08 8,45 15,465
	2	2,85 3,72 1,33	1,33 0,24 2,105	1,75 2,28 1,83	5,93 6,24 5,625
Total TIME		78,75 78,43 79,27	11,30 11,92 15,465	9,95 9,65 5,625	100,00 100,00 100,00

Test du Chi Deux :
Multinom. comp. = 2,77
Metheringham =10,58

FREQUENCES D'EXPOSITION

Nombre d'Expositions	Fréquences Observées	Fréquences estimées Multinom. comp.	Metheringham
0	71,42	70,16	69,89
1	11,66	12,72	16,10
2	12,03	14,18	7,97
3	3,14	0,66	4,21
4	1,75	2,28	1,83
Test d'ajustement du Chi Deux :		2,60	3,46
Couverture nette :	28,58	29,84	30,11

Ces résultats ne sont pas spécifiques à la paire de supports considérée. En fait, l'amélioration des estimations se vérifie pour tous les 57 plans d'insertion test de notre banque de données, ainsi que le montre la table suivante qui résume les performances du modèle Multinomial Composé.

Performances du Modèle Multinomial Composé

Couverture Nette

Erreur Absolue Moyenne	945 milliers d'individus
Pourcentage Moyen d'Erreur	2,1 %
Nombre de Surestimations	44 sur 57
Nombre d'estimations comprises dans l'intervalle de confiance à 95 %	49 sur 57
Ecart Type de l'Erreur	1568 milliers d'individus.

Fréquence d'Exposition

Chi Deux Moyen	69,68
Nombre de Modes correctement placés	57 sur 57
Nombre de Pics secondaires correctement placés (2 et 4 expositions)	114 sur 114

Le modèle Multinomial composé améliore donc nettement les performances du modèle de Metheringham sans pour autant nécessiter le recours à des données supplémentaires. Le pourcentage moyen d'erreur sur la couverture nette est diminué de moitié.

De plus, alors que le modèle de Metheringham surestime systématiquement la couverture nette, notre modèle produit 44 surestimations et 13 sous-estimations. L'erreur absolue moyenne est seulement de 945 milliers d'individus contre 2 483 milliers pour le modèle de Metheringham et 49 estimations sont comprises dans l'intervalle de confiance à 95 % contre 25 seulement pour le modèle de Metheringham. En résumé, le modèle Multinomial composé est supérieur au modèle de Metheringham dans 49 cas sur 57.

En ce qui concerne la distribution des occasions de voir, le Chi Deux moyen du modèle Multinomial composé est de 69,68 contre 68,77 pour le modèle de Metheringham. Notre modèle reproduit parfaitement la forme de la distribution des

occasions de voir puisque tous les pics sont correctement placés. Cependant, alors que le modèle de Metheringham efface les pics secondaires, notre modèle les accentue. Les fréquences paires sont surestimées et les fréquences impaires sont sous estimées alors que l'inverse se produit pour le modèle de Metheringham. En conclusion, nous avons montré qu'il est possible de construire un modèle à distribution exacte qui améliore nettement les performances du modèle de Metheringham sans pour autant nécessiter le recours à des données supplémentaires.

Nous allons tourner maintenant notre attention vers les divers modèles d'approximation utilisés aux U.S.A., en France et en Australie.

2 2 - Les Modèles d'approximation.

Les modèles d'approximation peuvent être construits selon trois approches.

- La première consiste à considérer deux supports, à les agréger ensemble de façon à constituer un nouveau support, puis à agréger ce nouveau support avec un troisième et ainsi de suite jusqu'à ce que tous les supports inclus dans le plan publicitaire aient été épuisés. Cette méthode, que nous intitulons agrégation séquentielle, est celle utilisée par le modèle de duplication de Starch et le modèle de duplication de Goodhart et Erhenberg que nous avons déjà étudié au chapître 6. Cette méthode présente un inconvénient important. Le résultat final dépend de l'ordre dans lequel les supports sont agrégés. Néanmoins, elle a permis de construire le meilleur modèle actuellement utilisé aux U.S.A., le modèle Modal, que nous allons étudier prochainement.

- La seconde approche, qui s'est développée spécialement en France, consiste à recourir à une formule de type Ad Hoc, ne reposant pas sur un modèle stochastique prédéterminé, mais permettant néanmoins d'obtenir une estimation rapide de la couverture nette. Nous étudierons le modèle d'exposition complet d'Hofmans et celui de Morgensztern. Nous présenterons aussi les quelques informations disponibles sur le modèle de casualité constante développé en Australie par Christopher Fry. Enfin, nous ajouterons à cette liste notre propre modèle d'approximation basé sur la combinaison des estimations du modèle de Metheringham avec celles du modèle Multinomial composé.

- La troisième approche consiste à recourir à la simulation des expositions d'un échantillon d'individus à qui l'on a attribué préalablement un vecteur de probabilités d'exposition à divers supports. Cette attribution des probabilités peut s'effectuer soit de manière empirique comme dans les modèles de Gensch et de Friedman, soit en recourant à une densité de probabilité comme dans le modèle de Greene et le modèle S.C.A.L. développé en France par la SEMA.

Modèle 8 Le Modèle Modal

Ce modèle a été construit à l'origine par David N. Sprague et il est commercialisé aux Etats Unis et au Canada par la société Interactive Market Systems (23). Il n'existe que trois publications décrivant le modèle Modal. Deux d'entre elles sont des communications présentées, à la 17e conférence annuelle

de l'"Advertising Research Foundation", par L.H. Libman (24) et par R.A. Metheringham (25). La troisième est un rapport publié par I.M.S. (26) qui compare les performances de deux versions, Modal I et Modal II, avec celles du modèle de Metheringham et du modèle des probabilités personnelles. Il ressort de ces documents que le modèle Modal est comparable au modèle Bêta matriciel sans toutefois nécessiter la tabulation de matrice d'exposition à m dimensions requise par le modèle Bêta Matriciel. La procédure suivie est la suivante :

- Premièrement, chaque support est étendu séparément, au nombre désiré de parutions, en recourant à la distribution Bêta Binomiale.

- Deuxièmement, les distributions des occasions de voir de chaque support sont agrégés séquentiellement dans un ordre prédéterminé.

La première étape a été discutée longuement au chapitre 5 et nous n'y reviendrons pas. La seconde étape est plus problématique car le modèle Modal ne nécessite pas la tabulation croisée de tous les supports, mais seulement la connaissance des duplications de chaque paire de supports. L'agrégation de deux supports est conduite sur la base des trois hypothèses suivantes :

a) - Chaque ligne et chaque colonne de la matrice d'exposition suit une distribution Bêta Binomiale.

b) - Le moment d'ordre un de chaque ligne ou colonne est une fonction linéaire du nombre d'expositions qui caractérise cette ligne ou colonne.

c) - Le moment d'ordre deux de chaque ligne ou colonne est une fonction quadratique du nombre d'expositions qui caractérise cette ligne ou colonne.

La première hypothèse est identique à celle du modèle Bêta Matriciel et nous avons montré qu'elle est justifiée, tant sur le plan pratique que sur le plan théorique, puisque les distributions conditionnelles du modèle Multinomial composé sont en fait des distributions Bêta Binomiales. La seconde et la troisième hypothèses caractérisent le modèle Modal. La seconde hypothèse a été soumise à plusieurs tests empiriques, par I.M.S. et Par Metheringham, pour divers plans publicitaires. Les tests ont montré que cette hypothèse est raisonnablement vérifiée par les données. De plus, si l'on admet la validité du modèle Multinomial composé, la

seconde hypothèse possède un fondement théorique puisque, dans ce modèle, les courbes de régressions sont linéaires ainsi que nous l'avons mentionné dans l'expression (46) de ce chapitre. La troisième hypothèse paraît avoir été choisie davantage pour des raisons pratiques que pour des raisons théoriques.

Puisqu'aucun autre détail n'a été publié sur le Modèle Modal à cause du secret commercial, tout ce qui suit ne représente rien de plus que les spéculations de l'auteur.

Tout d'abord, puisque Modal procède par agrégation séquentielle, le problème du choix de la meilleure séquence d'agrégation doit être résolu d'une façon ou d'une autre. Il se peut que le modèle ait recours à une méthode heuristique qui sélectionne l'ordre d'entrée des supports qui correspond au meilleur ajustement empirique de la distribution des occasions de voir. Ceci peut être accompli en prenant la valeur moyenne d'un échantillon de permutations systématiques de tous les supports.

Le second problème concerne l'estimation des duplications entre deux supports agrégés et un troisième support. Une solution consisterait à estimer les triplications en recourant à l'un des modèles décrits dans le chapitre 6. On obtiendrait alors la duplication nette entre les supports agrégés et le troisième support en soustrayant la triplication de la somme des duplications du troisième support avec chacun des deux supports agrégés.

Le dernier problème et le plus inquiétant, est celui de l'estimation des deux paramètres de la relation linéaire entre les moyennes des lignes (ou colonnes) et des trois paramètres de la relation quadratique entre les variances des lignes (ou colonnes). Pour deux supports, il faut estimer 4 paramètres linéaires et 6 paramètres quadratiques alors que l'on ne dispose que des audiences d'une et de deux parutions et de l'audience nette de la paire de supports, soit seulement 5 données pour 10 inconnues !

Il existe deux versions de Modal : Modal I et Modal II. La seule différence entre elles tient, semble-t-il, à la sélection de la meilleure séquence d'agrégation des supports. Les performances de Modal I sont très similaires à celles de notre propre modèle Multinomial composé. Nous ne présenterons donc que les résultats de Modal II qui est la version la plus perfectionnée.

La table suivante résume les performances de Modal II pour les 57 plans

publicitaires tests. Les résultats détaillés, plan par plan, sont présentés dans les tableaux de la page 514 à la page 519 .

Résumé des Performances de Modal II pour 57 plans publicitaires

Estimation de la couverture nette

Erreur absolue moyenne sur la couverture nette	=	299
Pourcentage moyen d'erreur sur la couverture nette	=	0,8 %
Nombre de sous-estimations de la couverture nette	=	51
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	57
Erreur type de prévision	=	425

Estimation de la distribution des occasions de voir

Chi Deux Moyen	=	3,38	
Nombre de concordances entre les modes	=	57	(sur 57)
Nombre de formes correctes	=	95	(sur 114)

Commentaires

Ces résultats montrent que Modal II est supérieur au modèle Multinomial composé. L'erreur absolue moyenne sur la couverture nette n'est que le tiers de l'erreur du modèle Multinomial composé. Le pourcentage moyen d'erreur est incroyablement faible, 0,8 % seulement contre 2,1 % pour le modèle Multinomial composé. Cependant, Modal II tend à sous estimer la couverture nette plus souvent qu'il ne la surestime. Néanmoins, les estimations de Modal II sont comprises dans l'intervalle de confiance à 95 % pour 51 plans contre 49 seulement pour le modèle Multinomial composé. En résumé, Modal II est supérieur à notre propre modèle à distribution exacte dans 47 cas sur 57.

En ce qui concerne la distribution des occasions de voir, le Chi Deux moyen est de 3,38 contre 69,7 pour le modèle Multinomial composé. L'ajustement est, en vérité, excellent. De plus, la forme de la distribution est parfaitement reproduite puisque Modal II, tout comme le modèle Multinomial composé, obtient une concordance quasi totale au niveau de 2 et 4 expositions.

Ces résultats indiquent clairement que Modal est l'un des meilleurs

modèles d'estimation actuellement disponibles. Cependant, puisque Modal recourt à une approximation dans l'estimation des paramètres, nous soupçonnons qu'il est soumis au phénomène de déclin de la couverture nette. Cet inconvénient est reconnu dans le rapport publié par I.M.S. (26), où il est indiqué que le phénomène est moins fréquent dans la version II que dans la version I.

En ce qui concerne le critère du volume des calculs, qui est directement lié au coût d'exploitation du modèle et à la possibilité de l'inclure dans un algorithme d'optimisation, le classement suivant est obtenu : Le modèle de Metheringham arrive en première position puisqu'il ne requiert que l'estimation de deux paramètres. Le modèle Multinomial composé occupe la deuxième position puisqu'il ne nécessite l'estimation que de $m + 1$ paramètres dans sa version univariée multinomiale composée.

Modal nécessite l'estimation de 10 paramètres pour chacune des m-1 agrégations séquentielles. Il occupe donc la troisième position, avant le modèle Bêta matriciel qui s'avère le plus lourd du point de vue du volume des calculs.

Nous allons maintenant étudier les modèles d'Hofmans et de Morgensztern qui caractérisent l'approche Ad Hoc du problème de l'estimation de la couverture nette et de la distribution des occasions de voir.

Modèle 9 Le Modèle d'Hofmans.

En dehors des modèles d'accumulation et de duplication déjà étudiés au chapitre 5 et 6, Hofmans (27) a présenté un modèle d'exposition complet de type Ad Hoc, obtenu par combinaison de son modèle d'accumulation et de son modèle de duplication. La table suivante présente les formules qui permettent de calculer l'accumulation, la duplication et la couverture nette d'un plan publicitaire.

Les formules d'Hofmans

Formule d'Accumulation

(48)

$$R_n = \frac{(nR_1)^2}{nR_1 + \frac{n(n-1)^a}{2} kd}$$

$$k = 2R_1/R_2$$

$$d = 2R_1 - R_2$$

$$a = \frac{\log[(3R_1 - R_3)R_2/(2R_1 - R_2)R_3]}{\log(2)}$$

Formule de Duplication

(49)

$$R_m = \frac{(\Sigma_{i=1}^{m} A_i)^2}{\Sigma_{i=1}^{m} A_i + \sum_{i \neq j} KD}$$

$$\sum_{i \neq j} KD = \Sigma_{i=1}^{m} \Sigma_{j=i+1}^{m-1} k_{ij} A_{ij}$$

$$k_{ij} = \frac{A_i + A_j}{A_i + A_j - A_{ij}}$$

Formule combinée

(50)

$$R_{nm} = \frac{(\Sigma_{i=1}^{m} n_i A_i)^2}{\Sigma_{i=1}^{m} n_i A_i + \sum_{i \neq j} KD + \sum_{i \neq j} kd}$$

$$\sum_{i \neq j} KD = \Sigma_{i=1}^{m} \Sigma_{j=i+1}^{m-1} [n_i n_j k_{ij} A_{ij}]$$

$$\sum_{i \neq j} kd = \Sigma_{i=1}^{m} \frac{n_i(n_i - 1)^a}{2} k_i d_i$$

Les formules d'Hofmans présentent quatre inconvénients majeurs :

- La formule d'accumulation surestime systématiquement la couverture nette cumulée de chacun des supports. De plus, elle requiert la connaissance de la couverture nette cumulée des trois premières parutions, qui est rarement disponible.

- La formule de duplication, quoique étant l'une des meilleures disponibles, tend à surestimer systématiquement la couverture nette.

- Aucune des formules d'Hofmans ne permet d'estimer la distribution des occasions de voir.

- La formule combinée est soumise au phénomène du déclin de la couverture nette (28).

Le premier de ces inconvénients peut être corrigé en modifiant la formule originale d'Hofmans. Nous allons montrer qu'il n'est pas nécessaire de connaître la couverture nette cumulée de la troisième parution. Cette modification nous permettra de tester la formule combinée d'Hofmans à l'aide de nos 57 plans publicitaires. Nous présenterons ensuite une autre modification de la formule d'Hofmans, introduite par Morgenztern.

Nous avons montré, dans le chapitre 5, que plusieurs modèles d'accumulation ont des performances supérieures à celles de la formule d'accumulation d'Hofmans, tout en ne nécessitant que la couverture nette cumulée des deux premières parutions. Nous avons aussi montré que, parmi ces modèles, la formule d'Agostini et le modèle Bêta Binomial conduisent à des estimations excellentes de la couverture nette cumulée. Il est donc possible de remplacer la formule d'accumulation d'Hofmans par l'un de ces deux modèles d'accumulation pour estimer l'audience cumulée à n parutions de chacun des supports inclus dans le plan publicitaire. Il suffit ensuite de calculer, à partir de ces estimations, la valeur de kd pour chacun des supports. Ceci est accompli en recourant à la formule suivante qui nous donne la valeur de kd en fonction de R_1, R_2 et R_n.

$$(51) \quad kd = \frac{2nR_1(nR_1 - R_n)}{nR_n\,(n-1)^a} \quad \text{avec } a = \frac{\log \dfrac{R_n(nR_1 - R_n)}{R_2(2R_1 - R_2)}}{\mathrm{Log}(n-1)}$$

La formule (51) a été obtenue en extrayant la valeur de kd de la formule (48). Etant donné que R_n peut être obtenue à partir de R_1 et de R_2 soit par la formule d'Agostini, soit par la formule Bêta Binomiale, il est clair que la connaissance de la couverture nette de la troisième parution n'est plus nécessaire.

L'exemple numérique suivant décrit en détail le calcul de la couverture nette du plan publicitaire n° 34 sur la base de la formule combinée d'Hofmans.

Exemple numérique :

Les seules données nécessaires sont l'audience des 2 premières parutions de chacun des supports et l'audience dupliquée de chaque paire de supports.

Magazines	Audiences 1 Parution	Audiences 2 Parutions	Audiences Dupliquées		$k_{ij}A_{ij}$	$n_i n_j$
B.H.G.	22 724	30 249	BHG-FH	6 124	7 329	4
F.H.	14 516	20 293	BHG-LHJ	6 578	7 898	4
L.H.J.	16 634	22 598	FH-LHJ	4 739	5 589	4
$\Sigma_{i=1}^{3} Ai$	= 53 874		$\sum_{i \neq j} KD$	=	83 264	

Puisque le plan publicitaire comporte 2 insertions dans chaque support, $n_1 = n_2 = n_3 = 2$, la somme des kd se réduit à : $\sum_{i \neq j} kd = \Sigma_{i=1}^{3} k_i d_i$

Magazine	k_i	d_i	$k_i d_i$
B.H.G.	1,5025	15 199	22 836
F.H.	1,431	8 739	12 502
L.H.J.	1,472	10 670	15 708
		$\sum_{i \neq j} kd$ =	51 046

Nous pouvons maintenant calculer la couverture nette du plan publicitaire à l'aide de la formule combinée (50) :

$$R = (2 * 53\,874)^2 / (2 * 53\,874 + 83\,264 + 51\,046) = 47\,962$$

Cette estimation est d'assez bonne qualité puisque la couverture nette réelle du plan n° 34 est de 49 239. Nous observons donc, pour ce plan, une sous estimation de 1 277, soit 2,6 % seulement. L'exemple numérique montre que la formule combinée d'Hofmans ne nécessite pas de lourds calculs. Toutefois, les performances d'ensemble, pour les 57 plans publicitaires, ne sont pas totalement satisfaisantes comme le montre la table suivante.

Résumé des Performances du Modèle d'Hofmans

Estimation de la couverture nette

Erreur absolue moyenne sur la couverture nette	=	1 107
Pourcentage moyen d'erreur sur la couverture nette	=	2,8 %
Nombre de sous-estimations de la couverture nette	=	35
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	47
Erreur type de prévision	=	1 881

Commentaires

L'erreur absolue moyenne est inférieure à celle du modèle de Metheringham, mais elle est supérieure à celle du modèle Multinomial composé. Le modèle d'Hofmans ne montre pas de tendance à la sur-estimation ou à la sous-estimation systématique puisque l'on observe 22 surestimations et 35 sous-estimations. Le pourcentage moyen d'erreur est de 2,8 % et les estimations n'excèdent l'intervalle de confiance que dans 10 cas sur 57. Ce résultat est nettement supérieur à celui du modèle de Metheringham, mais il est inférieur à celui du modèle Multinomial composé qui ne compte que 7 estimations excédant les bornes de l'intervalle de confiance. Une comparaison détaillée, plan par plan, est présentée dans les tables de la page 514 à la page 519 .

La formule d'Hofmans est nettement moins satisfaisante que le Modèle Modal qui conduit à des estimations en moyenne 5 fois plus précises et qui n'excèdent jamais les bornes de l'intervalle de confiance. Néanmoins, eu égard à la rapidité des calculs, on peut considérer que la formule d'Hofmans doit être

préférée à la formule de Metheringham, chaque fois qu'il n'est pas nécessaire d'estimer la distribution des occasions de voir. Bien entendu, l'absence de toute estimation de la distribution des occasions de voir constitue un handicap sévère pour la formule d'Hofmans. Il convient de remarquer qu'il n'est pas impossible de calculer la distribution des occasions de voir qui correspond à la couverture nette estimée par le modèle d'Hofmans. Pour cela, il faudrait calculer, par utilisations successives de la formule d'Hofmans, la matrice de couverture nette cumulée, puis convertir cette matrice en matrice d'exposition, et, enfin, en déduire la distribution des occasions de voir. Il ne paraît guère opportun de se livrer à ces lourds calculs supplémentaires puisque les estimations du modèle d'Hofmans sont, de toute manière, inférieures à celles des modèles Modal ou Multinomial composé qui permettent une estimation directe de la distribution des occasions de voir.

Nous allons maintenant étudier une modification de la formule combinée d'Hofmans qui a été proposée par Morgensztern.

Modèle 10 Le Modèle de Morgensztern

Ce modèle a été présenté par Morgensztern (28) en 1970. Mrogensztern introduit l'hypothèse suivante concernant la relation entre la duplication des audiences cumulées et la duplication des audiences moyennes d'une parution :

(52) $$\Delta_{ij} = A_{ij} \quad R_{ni} R_{nj} / (A_i B_j)$$

Δ_{ij} = Duplication entre les audiences cumulées des supports i et j.

A_{ij} = Duplication entre les audiences d'une parution des supports i et j.

R_{ni} = Couverture nette cumulée du support i après ni parutions.

R_{nj} = Couverture nette cumulée du support j après nj parutions.

A_i = Audience d'une parution du support i.

A_j = Audience d'une parution du support j.

Cette hypothèse peut être formulée comme suit : le rapport des duplications cumulées aux duplications moyennes est égal au produit des rapport des audiences cumulées aux audiences moyennes de chaque support. Nous allons tester cette hypothèse en utilisant les données de l'enquête Politz de 1958 qui fournit la couverture nette cumulée de diverses paires de magazines pour 2, 4 et 6

parutions de chaque magazine. Nous estimerons tout d'abord Δ_{ij} par application de la formule (52), puis nous calculerons la couverture nette estimée comme suit :

$$R_{ni,nj} = R_{ni} + R_{nj} - \Delta_{ij}. \tag{53}$$

Les tables suivantes comparent la couverture nette mesurée par Politz avec la couverture nette estimée, obtenue par la formule (53), pour dix paires de magazines.

Test de l'hypothèse de Duplication Cumulée de Morgensztern
(Enquête Politz 1958 - Milliers de foyers)

2 parutions dans chaque magazine

Paire de Magazines	Observée	Estimée	Paire de Magazines	Observée	Estimée
LIFE-POST	26,200	25,568	POST-R.D.	27,380	26,609
LIFE-LOOK	26,860	25,825	POST-L.H.J.	20,620	19,790
LIFE-R.D.	30,910	29,915	LOOK-R.D.	29,520	29,005
LIFE-L.H.J.	26,270	25,692	LOOK-L.H.J.	23,460	22,951
POST-LOOK	23,380	22,786	R.D.-L.H.J.	27,430	30,065

4 parutions dans chaque magazine

Paire de Magazines	Observée	Estimée	Paire de Magazines	Observée	Estimée
LIFE-POST	31,580	29,606	POST R.D.	32,290	30,026
LIFE-LOOK	32,050	28,794	POST-L.H.J.	25,710	23,125
LIFE-R.D.	35,720	32,973	LOOK-R.D.	34,600	32,927
LIFE-L.H.J.	31,550	29,757	LOOK-L.H.J.	28,970	27,247
POST-LOOK	29,090	27,040	R.D.-L.H.J.	32,080	29,977

6 parutions dans chaque magazine

Paire de Magazines	Observée	Estimée	Paire de Magazines	Observée	Estimée
LIFE-POST	34,250	31,158	POST-R.D.	34,830	31,330
LIFE-LOOK	34,610	29,486	POST L.H.J.	28,500	24,424
LIFE-R.D.	37,990	33,959	LOOK-R.D.	37,120	34,448
LIFE-L.H.J.	34,200	31,279	LOOK-L.H.J.	31,890	29,030
POST-LOOK	32,120	28,379	R.D.-L.H.J.	34,500	31,303

Ces tables montrent que la couverture nette estimée est systématiquement inférieure à la couverture nette réelle, à l'exception de la paire de magazines "Reader's Digest - Ladies Home Journal", dans le cas de 2 parutions. De plus, nous observons que le pourcentage moyen d'erreur s'accroît avec le nombre de parutions considérées. Il est de - 3,46 % pour 2 parutions, puis de - 7,12 %

pour 4 parutions et enfin de - 10,41 % pour 6 parutions. Il semble donc que l'hypothèse de Morgensztern ne soit pas vérifiée, du moins en ce qui concerne les audiences mesurées par Politz.

Pour les plans publicitaires comprenant plus de deux supports, Morgensztern propose la modification suivante de la formule d'Hofmans :

$$(54) \qquad R = \frac{(\Sigma_{i=1}^{m} R_{ni})^2}{\Sigma_{i=1}^{m} R_{ni} + \Sigma_{i=1}^{m} \Sigma_{j=i+1}^{m-1} k_{ij} A_{ij} R_{ni} R_{nj} / (A_i * A_j)}$$

$$\text{avec } k_{ij} = (A_i + A_j) / (A_i + A_j - A_{ij})$$

Deux commentaires peuvent être faits au sujet de cette nouvelle formule. Premièrement, il est clair qu'elle incorpore l'expression (52). Ceci revient à admettre que les duplications cumulées sont directement proportionnelles aux duplications moyennes. Deuxièmement, elle nécessite la connaissance des couvertures nettes cumulées à ni parutions pour chacun des supports. En conséquence, celles-ci devront avoir été estimées préalablement, pour le nombre désiré d'insertions, en recourant à l'un quelconque des modèles d'accumulation présentés dans le chapitre 5.

La formule (54) a été testée par Morgensztern, pour toute une variété de combinaisons de magazines hebdomadaires français, dont les audiences ont été mesurées par un panel de lecteurs pour une période de 8 semaines. Les résultats présentés dans l'article précité ont été très bons. La formule de Morgensztern a été testée également sur la base des données de l'enquête Politz de 1958. Nous présenterons ces résultats plus loin ainsi que ceux du test de la formule de Morgensztern pour nos 57 plans publicitaires de la banque de données I.M.S. Mais, auparavant, il convient de noter que tous ces tests partagent une limitation commune. Pour chacun d'eux, l'audience cumulée de chaque support, était mesurée par l'enquête, pour le nombre désiré de parutions. Ceci est évidemment le cas pour le panel Français, mais aussi pour l'enquête Politz. C'est aussi le cas pour les données d'I.M.S. puisqu'aucun plan publicitaire ne comporte plus de deux parutions dans un même support. En conséquence, les différences entre la couverture nette observée et la couverture nette estimée ne proviennent que de l'hypothèse de duplication incorporée dans la formule (54). Dans la pratique, nous ne savons pas comment les erreurs commises, en estimant préalablement les audiences cumulées de chaque support, se combinent avec

l'erreur due à l'hypothèse de duplication de la formule (54). L'exemple numérique suivant apporte une information à ce problème et présente en détail les calculs nécessaires à l'estimation de la couverture nette engendrée par 4 parutions des magazines Life, Post et Look.

Exemple numérique :

Les données d'audiences nécessaires au modèle de Morgensztern sont extraites de l'enquête Politz de 1958.

Magazines	1 Parution	4 Parutions	Paires	Audiences nettes	Duplications
LIFE	15 320	26 610	LIFE-POST	20 280	4 570
POST	9 530	17 930	LIFE-LOOK	20 930	6 080
LOOK	11 690	22 610	POST-LOOK	17 510	3 710

Somme des Audiences cumulées = 67 150

Nous calculons tout d'abord la somme, pour les 3 paires de magazine de la quantité : $k_{ij}\ A_{ij}\ R_{ni}\ R_{nj}\ /\ (A_i * A_j)$. Ceci est fait dans la table suivante :

Paire	A_i+A_j	k_{ij}	$k_{ij}A_{ij}$	$k_{ij}A_{ij}R_{ni}R_{nj}\ /\ (A_i * A_j)$
LIFE-POST	24 850	1,2253	5 599,8	18 300
LIFE-LOOK	27 010	1,2905	7 846,2	26 359
POST-LOOK	21 220	1,2119	4 496,1	16 361
			Somme =	61 020

La couverture nette estimée est maintenant obtenue en appliquant la formule (54) :

$$R = (67\ 150)^2\ /\ (67\ 150 + 61\ 020) = 35\ 181$$

La couverture nette réelle de plan est de 35 700 d'après l'enquête de Politz. Donc, dans ce cas, la formule de Morgensztern est extrêmement précise avec une infime surestimation de 0,3 %. Cependant, qu'en eut-il été si nous n'avions pas connu l'audience cumulée à 4 parutions de chaque support, mais seulement l'audience des 2 premières parutions ? Dans ce cas, utilisant le modèle Bêta Binomial

d'accumulation, nous aurions trouvé :

Magazines	Audience estimée de 4 parutions	$k_{ij}A_{ij}R_{ni}R_{nj} / (A_i * A_j)$
LIFE	26 240	17 606
POST	17 494	25 590
LOOK	22 260	15 716
Somme =	65 994	58 912

$$R = (65\ 994)^2 / (65\ 994 + 58\ 912) = 34\ 867$$

Maintenant, la formule de Morgensztern sous-estime très légèrement la couverture nette, d'environ 0,6 %, ce qui est toujours un excellent résultat.

La table suivante présente un échantillon de résultats tirés de l'enquête Politz de 1958. Le pourcentage moyen d'erreur est de 1,2 % pour 4 parutions de 3 magazines, 0,8 % pour 3 parutions de 4 magazines et de 0,6 % pour 2 parutions de 3 magazines.

Test de la Formule de Morgensztern d'après les données de l'enquête Politz de 1958

(Milliers de foyers)

4 Parutions de 3 Supports

Combinaison de Magazines	Observé	Estimé	Combinaison de Magazines	Observé	Estimé
LIFE + POST + LOOK	35,070	35,181	LIFE + R.D. + L.H.J.	37,500	37,853
LIFE + POST + R.D.	37,600	37,993	POST + LOOK + R.D.	36,760	36,961
LIFE + POST + L.H.J.	34,460	34,060	POST + LOOK + L.H.J.	32,780	32,067
LIFE + LOOK + R.D.	38,110	38,729	POST + R.D. + L.H.J.	34,880	34,199
LIFE + LOOK + L.H.J.	35,120	35,296	LOOK + R.D. + L.H.J.	36,600	37,134

3 Parutions de 4 Supports

Combinaison de Magazines	Observé	Estimé	Combinaison de Magazines	Observé	Estimé
LIFE + POST + LOOK + R.D.	37,920	38,467	LIFE + LOOK + R.D. + L.H.J.	37,860	38,547
LIFE + POST + LOOK + L.H.J.	35,400	35,325	POST + LOOK + R.D. + L.H.J.	36,510	36,556
LIFE + POST + R.D. + L.H.J.	37,310	37,399			

2 Parutions de 3 Supports

Combinaison de Magazines	Observé	Estimé	Combinaison de Magazines	Observé	Estimé
LIFE + POST + LOOK	30,310	30,317	LIFE + R.D. + L.H.J.	33,210	33,392
LIFE + POST + R.D.	33,260	33,333	POST + LOOK + R.D.	32,120	32,247
LIFE + POST + L.H.J.	29,540	29,337	POST + LOOK + L.H.J.	27,470	27,161
LIEF + LOOK + R.D.	33,920	34,180	POST + R.D. + L.H.J.	30,290	29,970
LIFE + LOOK + L.H.J.	30,370	30,468	LOOK + R.D. + L.H.J.	32,070	32,334

Cette table montre que le modèle de Morgensztern achève de bons résultats avec les données de l'enquête Politz 1958. Toutefois, les résultats sont nettement moins bons avec les données I.M.S. ainsi que le montre la table suivante qui résume les performances du modèle de Morgensztern pour 57 plans publicitaires.

Résumé des Performances du Modèle de Morgensztern

Estimation de la couverture nette

Erreur absolue moyenne sur la couverture nette	=	1 106
Pourcentage moyen d'erreur sur la couverture nette	=	2,80 %
Nombre de sous-estimations de la couverture nette	=	36
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	46
Erreur type de prévision	=	1 905

Commentaires

Dans l'ensemble, les résultats du modèle de Morgensztern sont très similaires à ceux du modèle d'Hofmans. Les rares différences entre les deux groupes d'estimations sont légèrement en faveur du modèle d'Hofmans. Les résultats détaillés, plan par plan, sont présentés dans les tables de la page 514 à la page 519 .

En résumé, la principale supériorité du modèle de Morgensztern par rapport au modèle d'Hofmans tient à ce qu'il est exempt du phénomène de déclin de la couverture nette.

Pour clore cette étude des modèles d'exposition complets de type Ad Hoc, on peut affirmer que ceux-ci sont supérieurs au modèle de Metheringham bien qu'ils soient inférieurs au modèle Modal qui utilise l'approche séquentielle et au modèle Multinomial composé qui utilise une distribution exacte. Ils sont aussi inférieurs au modèle de casualité constante de Christofer Fry que nous allons étudier maintenant.

Modèle 11 Le Modèle de Casualité Constante

Ce modèle a été proposé par Christopher Fry, un spécialiste Australien de la planification des supports publicitaires. Un bref compte rendu décrivant le concept de casualité de lecture a été publié dans le numéro d'avril 1973 de la revue Feedback (29). Il n'existe aucun autre document publié sur ce modèle. Tout ce qui suit est basé sur l'article précité et sur un échange de correspondance avec C. Fry (30).

La casualité de lecture d'un support est un coefficient qui mesure le taux d'accumulation des audiences au cours des parutions successives. Ce coefficient est en relation directe avec les deux paramètres du modèle d'accumulation Bêta Binomial. Le coefficient de casualité est le rapport de l'audience exclusive de la seconde parution à l'audience exclusive qui serait obtenue sous l'hypothèse d'une accumulation aléatoire.

Si nous appelons R_1 et R_2 l'audience cumulée des deux premières parutions, l'audience exclusive de la seconde parution est égale à $R_2 - R_1$. Toutefois, si l'accumulation était purement aléatoire, l'audience exclusive de la seconde parution serait égale à $R_1 - (R_1)^2$. Le rapport de ces deux audiences exclusives définit le coefficient de casualité.

$$(55) \qquad c = (R_2 - R_1) \;/\; \left[R_1 - (R_1)^2 \right]$$

Si les audiences cumulées sont exprimées en valeur absolue et non pas en pourcentage de la population cible, le coefficient de casualité est calculé comme suit :

$$(56) \qquad c = (R_2 - R_1) \;/\left[R_1 - (R_1)^2 \;/\; POP \right]$$

Puisque les audiences ne s'accumulent pas au hasard, mais en fait à un taux beaucoup plus faible, le coefficient c mesure la déviation, par rapport à l'accumulation aléatoire. Cette déviation est supposée demeurer constante quel que soit le nombre de parutions considérées, comme l'indique le nom du modèle.

Par exemple, l'étude Politz de 1958 montre qu'une parution de Life était lue par 32,2 % des foyers et deux parutions par 42,6 %. L'audience ajoutée par la seconde parution était de 11,4 %. Si l'audience s'était accumulée au hasard, la seconde parution aurait apporté $0{,}312 - (0{,}312)^2$, soit 21,47 % de foyers supplémentaires. Par conséquent, le coefficient de casualité de Life

était de : 0,114 / 0,2147 = 0,531. Puisque le coefficient de casualité ne dépend que de l'audience cumulée des deux premières parutions, il doit nécessairement être relié au taux de rotation des audiences, t, et aux paramètres A et B du modèle Bêta Binomial. Le taux de rotation des audiences est le rapport de l'audience exclusive de la seconde parution à l'audience de la première parution. Nous avons donc :

$$(57) \quad t = (R_2 - R_1)/R_1 \qquad t = c\,(1 - R_1) \qquad c = t\,/\,(1 - R_1)$$

Nous pouvons aussi exprimer c en fonction des paramètres A et B :

$$(58) \quad c = (A + B)\,/\,(A + B + 1);\ A = cR_1/(1-c)\ ;\ B = c(1 - R_1)\,/\,(1 - c)$$

En fait, le coefficient de casualité n'est rien d'autre qu'une reformulation du modèle Bêta Binomial obtenue en remplaçant les paramètres A et B par R_1 et c. Toutefois, cette nouvelle formulation donne un aperçu différent du phénomène d'accumulation des audiences. Elle montre que le taux d'accumulation dépend du degré de casualité dans les lectures. S'il n'y avait pas d'habitudes de lecture ni de fidélité des audiences, l'accumulation serait maximum. Plus un support possède de lecteurs fidèles, plus faible sera le coefficient de casualité et plus faible sera le taux d'accumulation.

La contribution de Fry ne s'arrête pas là. Il a aussi mis au point une méthode, qui reste inédite, permettant d'étendre le concept de casualité au phénomène de duplication des audiences. Tout ce qui suit sur ce sujet ne représente que les extrapolations de l'auteur et ne saurait en aucun cas être tenu pour la méthode originale de Fry qui reste secrète.

Au premier abord, il semble aisé de transposer le concept de casualité à la duplication des audiences. Considérons deux supports. L'accroissement d'audience dû à l'addition d'un deuxième support est analogue à l'accroissement d'audience dû à une deuxième parution. On pourrait donc définir la casualité de lecture dans le cas de supports de la façon suivante :

	En Pourcentage de la Population	En Valeurs Absolues
(59)	$c_{A+B} = \dfrac{R_{AB} - R_A}{R_A - (R_A)^2}$	$c_{A+B} = \dfrac{R_{AB} - R_A}{R_A - (R_A)^2/POP}$

avec R_{AB} = Audience nette de la paire de supports A, B.

Toutefois, si, au lieu d'ajouter le support B au support A, nous ajoutons le support A au support B, le coefficient de casualité devient :

En Pourcentage de la Population — En Valeurs Absolues

$$(60) \quad c_{B+A} = \frac{R_{AB} - R_B}{R_B - (R_B)^2} \qquad c_{B+A} = \frac{R_{AB} - R_B}{R_B - (R_B)^2/POP}$$

On observe que le coefficient à change. Il est donc clair que cette première tentative de définition doit être rejetée puisque la valeur du coefficient de casualité dépend de l'ordre dans lequel les supports sont considérés comme le montre l'exemple numérique suivant tiré de l'enquête Simmons de 1973.

Magazines	Audiences	Audience netta de la paire	Audience dupliquée	Population
1 - B.H.G.	22 724	31 116	6 124	130 326
2 - F.H.	14 516			

$$c_{1+2} = (31\,116 - 22\,724) / (22\,724 - 3\,962) = 0{,}447$$

$$c_{2+1} = (31\,116 - 14\,516) / (14\,516 - 1\,617) = 1{,}287$$

Afin de supprimer cette difficulté, la solution la plus simple consiste à remplacer les audiences de chaque support par leur moyenne. Cette solution peut être justifiée en recourant à la distribution des occasions de voir générée par nos deux supports.

	Matrice d'Exposition			Distribution des Occasions de Voir		
	0 BHG	1 BHG	Total	0	1	2
0 F.H.	99 210	16 600	115 810	99 210	24 992	6 124
1 F.H.	8 392	6 124	14 516			
Total	107 602	22 724	133 326			

Posons nous la question suivante. Quelle est l'audience du support unique conduisant à la même distribution des occasions de voir que la paire B.H.G.-F.H. ? La réponse est obtenue en utilisant la relation de récurrence (33) de la page 194 du

chapitre 5.

$$R_1^* = \frac{1}{2} E_1^2 + E_2^2 = 18\ 620$$

Or, R_1^* n'est rien d'autre que la moyenne arithmétique des audiences des deux supports.

$$R_1^* = \bar{R}_1 = \frac{1}{2} (R_A + R_B)$$

En tenant compte de cette modification, le coefficient de casualité entre supports peut être défini comme suit :

En Pourcentage de la Population	En Valeurs Absolues
$c_{ij} = \dfrac{R_{ij} - \bar{R}_1}{\bar{R}_1 - (\bar{R}_1)^2}$	$c_{ij} = \dfrac{R_{ij} - \bar{R}1}{\bar{R}_1 - (\bar{R}_1)^2/POP}$

L'expression (61) rend le coefficient de casualité indépendant de l'ordre d'entrée des supports. En utilisant B.H.6 et F.H. on obtient par exemple :

$$c = (31\ 116 - 18\ 620) / (18\ 620 - 2\ 660) = 0{,}783$$

Il convient aussi de noter que l'expression (61) est équivalente à l'ajustement d'une loi Bêta Binomiale à la distribution des occasions de voir de la paire de supports. Toutefois, un problème se pose lorsque 3 supports en plus sont considérés. Dans ce cas, il est impossible de reconstituer directement la distribution des occasions de voir. Il nous faut d'abord estimer les triplications et les recouvrements d'ordre supérieur à deux. Nous ne savons pas comment ce problème a été résolu par Fry. Cependant, puisque Fry précise qu'il n'utilise qu'un seul coefficient pour représenter la casualité de lecture entre les supports, il doit nécessairement recourir à une moyenne pondérée des coefficients de casualité de chacune des paires de supports. Si c'est le cas, sa méthode conduira dans certains cas à l'apparition du phénomène de déclin de la couverture nette.

Finalement, la méthode de Fry transpose le problème de l'estimation de la couverture nette en un problème d'estimation du coefficient de casualité. Ceci ne pose pas de difficultés pour la casualité de chacun des supports pris séparément et peut être obtenue par l'une quelconque des méthodes d'estimations

présentées au chapitre 5. Par contre, on ne voit guère comment la casualité entre les supports peut être estimée de façon satisfaisante, puisque nous avons montré que la distribution Bêta Binomiale est incapable de reproduire les pics secondaires qui caractérisent la distribution des occasions de voir générée par une combinaison de supports.

Nous ne savons pas non plus comment la distribution des occasions de voir est estimée à partir des deux coefficients de casualité ! Néanmoins, nous pouvons montrer comment utiliser le concept de casualité pour estimer la couverture nette d'un plan publicitaire faisant appel à deux supports.

Exemple numérique

Reprenons notre exemple habituel des magazines Time et Newsweek et supposons que nous ne connaissions que l'audience des 2 premières parutions de chaque support ainsi que l'audience nette de la paire de supports.

Magazines	1 Parution	2 Parutions	Audience nette de la Paire de Supports	Population
Time	20 832	28 378	29 216	133 520
Newsweek	13 984	20 049		
Moyenne	17 408			

Nous pouvons estimer la casualité entre les supports par la formule (61) =

$$c = (29\,216 - 17\,408) / (17\,408 - 2\,270) = 0{,}78$$

Afin d'estimer la couverture nette de 2 insertions dans chaque support, nous devons d'abord calculer la couverture nette sous l'hypothèse de duplication aléatoire, puis appliquer ensuite le coefficient de casualité. Si les audiences de Time et de Newsweek étaient indépendantes, le nombre d'individus non exposés aux deux supports serait égal à :

$$119\,536 * 112\,688 / 133\,520 = 100\,886$$

Sous la même hypothèse, le nombre d'individus non exposés à deux parutions de chaque support serait égal à :

$$113\,471 * 105\,142 / 133\,520 = 89\,354$$

L'accroissement de couverture nette, sous l'hypothèse de duplication aléatoire est donc égal à 11 532. Mais nous savons que seulement 78 % de cet accroissement sera observé puisque le coefficient de casualité est de 0,78. Donc, l'accroissement estimé de couverture nette sera de 8 995 seulement. Puisque la couverture nette de la paire de supports est de 29 216, la couverture nette estimée de 2 insertions dans chaque support est de :

$$29\ 216 + 8\ 995 = 38\ 211$$

En fait, la couverture nette observée est de 38 165, donc la méthode de casualité constante fournit dans ce cas une estimation extrêmement précise puisque l'erreur n'est que de 0,12 %.

Le modèle de Fry a été testé contre 45 plans publicitaires parmi les 57 plans de notre banque de données. Quatre versions du modèle de Fry ont été testées, mais nous ne présenterons ici que les meilleurs résultats qui ont été obtenus par la version 4.

Résumé des Performances du Modèle de Casualité Constante pour 45 plans publicitaires

Estimation de la couverture nette		
Erreur moyenne absolue sur la couverture nette	=	328
Pourcentage moyen d'erreur sur la couverture nette	=	0,88 %
Nombre de surestimations de la couverture nette	=	32
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	45
Erreur type de prévision	=	450
Estimation de la distribution des occasions de voir		
Chi Deux moyen	=	10,15
Nombre de Modes correctement placés	=	45 sur 45
Nombre de pics secondaires correctement placés (2 et 4 expositions)	=	86 sur 90

Commentaires

La version 4 du modèle de casualité constante obtient des résultats presque identiques à ceux obtenus par Modal II. L'erreur absolue moyenne est de 328 contre 299 pour Modal II et le pourcentage moyen d'erreur est de 0,88 %

contre 0,8 %. Les deux méthodes produisent des estimations sans biais systématique, les surestimations étant légèrement plus fréquentes pour le modèle de casualité constante que pour le modèle Modal. Dans les deux cas, toutes les estimations sont comprises dans l'intervalle de confiance à 95 % autour des données de l'enquête. Finalement, une comparaison, plan par plan, montre que le modèle de casualité constante est supérieur à Modal pour 15 plans, tandis que Modal est supérieur pour les 30 plans restants. En conclusion, il est presque impossible de départager les deux méthodes sur le plan de l'estimation de la couverture nette.

En ce qui concerne l'estimation de la distribution des occasions de voir, Modal est assez nettement supérieur au Modèle de casualité constante. En effet, le Chi Deux Moyen de Modal n'est que de 3,28 contre 10,15 pour le modèle de Fry. Toutefois, les deux modèles reproduisent presque parfaitement la forme de la distribution des occasions de voir. Tous les modes sont correctement reproduits, tandis que 93 pics secondaires sur 114 sont reproduits par Modal contre 86 sur 90 pour le modèle de casualité constante.

Etant donné l'excellente qualité des résultats obtenus par les deux modèles, il semble difficile que des améliorations ultérieures puissent être apportées sans recourir à des données supplémentaires. Néanmoins, puisque ces deux modèles sont couverts par le secret commercial, nous allons proposer notre propre modèle d'approximation. Bien qu'il ne conduise pas à des estimations aussi bonnes que celles des deux modèles précédents, il présente néanmoins l'avantage de nécessiter des calculs moins volumineux.

Les six tables suivantes présentent une comparaison détaillée, plan publicitaire par plan publicitaire, des performances des 7 modèles d'exposition complets à population hétérogène que nous venons d'étudier.

Comparaison des Performances de 7 Modèles d'Exposition à Population Hétérogène

Estimation de la Couverture Nette

Plan Publicitaire	Observé	Metheringham	Multinomial Composé	Modal II	Hofmans	Morgensztern	Casualité Constante
1	95 511	98 046*	96 147	95 191	96 890	96 713	95 499
2	96 644	100 247*	98 058*	97 357	99 630*	99 653*	97 348
3	94 547	99 864*	96 119	96 113	98 327*	98 304*	95 677
4	64 543	68 439*	61 527*	64 314	61 007*	61 295*	64 080
5	57 047	59 868	64 281*	56 974	65 805*	66 063*	57 242
6	51 838	53 977	56 786*	51 776	57 940*	58 215*	51 941
7	31 645	34 470*	31 785	31 675	31 514	31 499	31 596
8	28 843	30 703	29 383	28 904	28 346	28 367	28 847
9	30 800	32 924	31 276	30 853	30 554	30 595	30 828
10	11 527	11 761	11 759	11 235	10 806	10 811	11 524
11	16 523	17 489	16 987	16 549	16 351	16 330	16 652
12	15 425	16 190	15 574	15 344	15 178	15 183	15 353
13	66 141	72 021*	68 106*	66 490	66 861	66 430	66 597
14	62 399	67 260*	64 292	63 033	63 336	62 944	63 268
15	65 675	70 447*	68 305*	65 289	65 907	65 035	65 892
16	26 106	27 738*	27 199	26 621	26 359	26 459	26 818
17	23 980	25 304	24 209	24 091	23 973	23 963	24 121
18	14 548	15 072	15 181	14 516	14 062	14 054	14 875
19	13 598	14 389	13 938	13 571	13 242	13 225	13 672
20	56 531	58 380	55 686	56 106	54 657	54 573	57 057
21	50 095	52 358*	48 288	49 795	49 563	49 496	49 819
22	51 426	55 009*	52 416	51 801	50 907	50 911	52 141
23	60 542	63 883	60 796	60 735	60 055	60 042	60 743
24	37 766	40 803	37 265	37 699	37 623	37 618	37 579
25	82 880	89 759*	82 669	83 671	84 949	84 927	83 207

Estimation en dehors d'un intervalle de confiance à 95 % autour des données de l'enquête

Comparaison des Performances de 7 modèles d'Exposition à Population Hétérogène

Estimation de la Couverture Nette (suite)

Plan Publicitaire	Observé	Metheringham	Multinomial Composé	Modal II	Hofmans	Morgenstern	Casualité Constante
26	84 402	89 869*	84 146	84 850	86 351	86 401	84 434
27	92 627	97 677*	93 467	93 592	95 426*	95 418*	93 168
28	77 643	82 043*	78 566	77 896	78 353	78 574	77 804
29	73 215	77 885*	75 588*	73 001	73 680	72 981	73 485
30	50 600	53 535*	52 008	51 073	50 329	50 241	51 273
31	9 349	9 540	11 740	9 161	8 241*	8 291*	9 383
32	13 731	14 217	14 197	13 615	13 271	13 271	13 895
33	21 487	22 567	23 147	22 078	20 812	20 862	22 860
34	49 239	50 838	50 332	49 096	47 962	47 901	49 824
35	50 460	51 318	50 768	49 546	48 379*	48 325*	50 268
36	58 819	61 054	58 558	58 938	58 388	58 520	59 196
37	22 693	24 509	22 949	22 748	22 523	22 597	22 640
38	89 625	93 722*	89 561	90 179	91 297	91 361	89 738
39	41 935	46 190*	42 792	42 003	42 120	42 144	41 943
40	67 547	70 108	68 067	67 957	67 534	67 737	68 299
41	44 490	46 232	44 193	43 885	42 138	43 371	44 645
42	32 436	33 967	32 987	31 939	30 842	30 857	33 289
43	39 172	41 332	39 429	38 753	38 276	38 280	38 964
44	50 077	51 379	48 150	49 311	48 820	48 783	49 498
45	74 913	79 259*	74 794	75 162	75 384	75 250	75 390
46	49 047	50 974*	48 951	48 797	49 836	49 749	—
47	38 832	39 325	39 035	38 460	37 101*	37 051*	—
48	17 302	18 539	18 330	17 229	16 999	16 884	—
49	24 097	24 807*	23 984	23 988	24 749*	24 748*	—
50	20 238	20 502	20 359	20 131	19 619	19 602	—

* Estimation en dehors d'un intervalle de confiance à 95 % autour des données de l'enquête

Comparaison des Performances de 7 Modèles d'Exposition à Population Hétérogène

Estimation de la Couverture Nette

(Suite)

Plan Publicitaire	Observé	Metheringham	Multinomial composé	Modal II	Hofmans	Morgenstern	Casualité Constante
51	10 747	11 524	11,244	10 748	10 667	10 585	--
52	10 858	11 220	11.014	10 779	11 180	11.126	--
53	9 519	9 593	9 610	9 384	9,233	9 209*	--
54	7.242	7 464	7 333	7 173	7,156	7 126	--
55	3 114	3 185	3,264	3 108	3 229	3,257	--
56	616	625	665	604	568	574	--
57	2 638	2 757	2 662	2 691	2 712	2 726	--
Erreur absolue moyenne		2 483	945	299	1,107	1.106	406

Estimation en dehors d'un intervalle de confiance à 95 % autour des données de l'enquête.

Comparaison des Performances de 4 Modèles d'Exposition et Population Hétérogène

Estimation de la Distribution des Occasions de Voir

Valeurs du Test d'Ajustement du Chi Deux (Base 13 000)

Plan Publicitaire	Metheringham	Multinomial Composé	Modal II	Casualité Constante	Probabilités Personnelles
1	22,50	173,81	0,77	13,00	-
2	36,20	153,92	0,28	5,56	-
3	110,50	179,53	1,36	8,53	6,31
4	90,70	92,30	0,76	17,63	10,47
5	165,50	123,37	1,21	7,14	-
6	103,20	85,15	2,55	3,41	-
7	76,30	21,06	0,78	9,22	11,79
8	39,70	23,40	1,29	18,34	-
9	49,40	19,89	1,43	14,34	-
10	41,70	17,68	2,29	8,06	-
11	14,60	13,00	0,84	4,96	21,77
12	21,4	8,19	3,61	5,08	-
13	183,4	50,70	4,09	26,17	-
14	109,7	60,84	4,20	18,04	-
15	117,3	52,65	11,02	43,35	2,52
16	26,52	29,12	5,02	6,70	-
17	24,44	11,96	0,47	0,48	-
18	7,54	12,74	1,47	2,75	17,68
19	12,22	14,95	2,43	7,34	-
20	24,70	52,13	7,41	3,79	56,18
21	55,90	26,26	4,56	12,56	-
22	78,00	55,12	2,59	12,57	44,32

Comparaison des Performances de 4 Modèles d'Exposition et Population Hétérogène

Estimation de la Distribution des Occasions de Voir

Valeurs du Test d'Ajustement du Chi Deux (Base 13 000) (suite)

Plan Publicitaire	Metheringham	Multinomial Composé	Modal II	Casualité Constante	Probabilités Personnelles
23	67,99	67,73	1,49	8,78	-
24	109,07	20,54	0,39	10,75	-
25	203,45	126,10	8,27	5,81	1,08
26	163,15	118,30	2,84	10,83	6,42
27	116,35	164,97	7,31	9,57	7,24
28	84,24	106,21	3,55	15,20	-
29	102,49	70,98	7,56	3,84	7,82
30	56,29	55,12	5,73	6,07	-
31	2,73	33,15	1,27	3,12	13,27
32	5,85	12,61	0,31	1,34	-
33	19,76	36,66	13,32	13,57	-
34	37,18	59,93	1,50	2,32	-
35	28,73	66,69	2,80	6,35	-
36	28,73	60,58	5,32	3,09	-
37	49,79	24,96	3,91	18,05	35,41
38	121,81	118,30	4,41	6,06	-
39	123,37	34,71	2,08	1,20	-
40	42,90	68,25	0,87	4,38	2,66
41	39,00	61,62	7,70	19,54	-
42	80,08	65,13	3,89	36,27	-
43	52,78	39,78	0,46	7,42	-
44	69,16	25,22	6,11	6,07	-

Comparaison des Performances de 4 Modèles d'Exposition et Population Hétérogène

Estimation de la Distribution des Occasions de Voir

Valeurs du Test d'Ajustement du Chi Deux (Base 13 000) (suite)

Plan Publicitaire	Metheringham	Multinomial Composé	Modal II	Casualité Constante	Probabilités Personnelles
45	80,21	70,46	13,44	8,26	4,71
46	77,87	137,41	5,80	-	-
47	96,72	132,60	3,11	-	-
48	54,86	41,86	2,50	-	-
49	76,96	148,33	3,41	-	-
50	30,42	142,35	2,16	-	-
51	80,08	59,54	1,88	-	-
52	71,89	139,75	1,86	-	-
53	48,71	119,08	2,59	-	-
54	44,20	99,58	1,33	-	10,12
55	127,92	99,97	1,54	-	2,46
56	12,87	31,85	0,26	-	-
57	195,39	32,28	1,29	-	7,22
Moyenne	68,77	69,68	3,38	10,15	13,89

Modèle 12 Le Modèle Metheringham - Multinomial Composé

Ce modèle est obtenu par la simple moyenne arithmétique des distributions des occasions de voir estimées par le modèle de Metheringham et par le modèle Multinomial composé. Ces deux modèles à distribution exacte donnent des résultats opposés, comme nous l'avons déjà remarqué. Metheringham surestime les fréquences paires, tandis que le modèle Multinomial composé les sous-estime. De même, Metheringham sous-estime les fréquences impaires tandis que le modèle Multinomial composé les surestime. En prenant la moyenne des deux estimations, on obtient donc une meilleure estimation de la distribution des occasions de voir. Néanmoins, étant donné que les deux modèles surestiment la couverture nette, la méthode conduira aussi à une surestimation de la couverture nette. Les gains obtenus ne seront donc sensibles qu'en ce qui concerne l'ajustement de la distribution des occasions de voir.

La table suivante présente les performances obtenues par le modèle Metheringham-Multinomial composé.

Résumé des Performances du Modèle Metheringham-Multinomial Composé pour 57 Plans Publicitaires

Estimation de la couverture nette

Erreur absolue moyenne sur la couverture nette	=	979
Pourcentage moyen d'erreur sur la couverture nette	=	2,73 %
Nombre de surestimations de la couverture nette	=	36
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	50
Erreur type de prévision	=	1 684

Estimation de la distribution des occasions de voir

Chi Deux moyen	=	19,1
Nombre de Modes correctement placés	=	57 sur 57
Nombre de pics secondaires correctement placés (2 et 4 expositions)	=	88 sur 114

Commentaires :

Les performances de ce modèle sont clairement inférieures à celles du modèle Modal ou du modèle de casualité constante. Toutefois, elles représentent une grande amélioration par rapport aux estimations du modèle de Metheringham et du modèle Multinomial composé. L'erreur absolue moyenne est deux fois et demie moindre que celle du modèle de Metheringham et le pourcentage moyen d'erreur est deux fois moindre. De plus, la tendance à la surestimation systématique est nettement moins prononcée. Sept estimations seulement sont à l'extérieur de l'intervalle de confiance à 95 % autour des données de l'enquête contre 23 pour le modèle de Metheringham. L'erreur type de prévision est moitié moindre que celle du modèle de Metheringham.

Toutefois, l'amélioration la plus évidente se produit au niveau de l'estimation de la distribution des occasions de voir. Le Chi Deux moyen est quatre fois moindre que celui du modèle de Metheringham ou du modèle Multinomial composé. La forme de la distribution des occasions de voir est reproduite de façon satisfaisante avec 88 concordances sur 114 au niveau des pics secondaires.

Nous avons maintenant terminé notre étude des modèles d'exposition completsde type stochastique. Nous allons maintenant examiner les performances des modèles de simulation.

o o
o

3 - Les Modèles de Simulation.

Les Partisans des Modèles de simulation ne se sont pas montrés extrêmement enclins à vérifier, de façon empirique, la validité des estimations de leurs modèles, à l'exception toutefois du modèle des probabilités personnelles de Greene et du modèle de simulation S.C.A.L., mis au point par des chercheurs Français.

Nous avons déjà noté que le coeur des méthodes de simulation réside dans l'affectation des probabilités d'exposition aux individus simulés. Nous avons aussi démontré que l'affectation actuarielle et l'affectation empirique des probabilités conduisaient nécessairement à des estimations peu précises de la couverture nette et de la distribution des occasions de voir puisqu'elles échouent sur les critères de reproduction des données originales et d'unicité de la solution. En conséquence, nous ne discuterons pas ici du Modèle de Gensch, "Ad-ME-SIM" (31), ni du modèle d'Aaker, "ADMOD" (32) qui utilisent, tous les deux, l'affectation actuarielle des probabilités et qui n'ont pas été soumis à des validations empiriques.

D'une manière générale, il semble que les partisans des modèles de simulation partagent l'opinion suivante exprimée par Friedman (33) : "Si la couverture nette et la distribution des occasions de voir estimée par le modèle de simulation sont à la fois plausibles et logiques, c'est-à-dire, si elles approximent les estimations obtenues par un spécialiste compétent utilisant les meilleures méthodes disponibles, alors le modèle de simulation peut être utilisé pour prendre des décisions concernant le choix des supports".

Nous ne partageons pas ce point de vue. Les estimations d'un expert ne constituent pas, en soi, un critère de validité. S'il en était ainsi, pourquoi prendre la peine de construire des modèles d'exposition ? Nous admettons volontiers que les données d'audience ne sont pas parfaites et nous avons discuté en détail leurs diverses limitations dans le chapitre 4. Nous admettons aussi que les classifications croisées des résultats d'enquêtes ou de panel coûtent fort cher. Néanmoins, aucun modèle ne peut prétendre être empiriquement validé tant que ses estimations n'ont pas été comparées avec des données d'enquête. Deux tests de ce genre ont été réalisés et ce n'est pas un hasard si les deux modèles testés sont aussi ceux qui vérifient les critères théoriques de reproduction des données originales et d'unicité de la solution. Le premier test a été conduit avec la banque de données I.M.S., le second test avec les données d'un panel de

lecture Français. Nous présenterons aussi les résultats d'un troisième test de moindre envergure réalisé par les constructeurs du modèle de simulation Anglais C.A.M.

31 - Le modèle des probabilités personnelles.

Nous avons déjà montré au chapître 5 que cette méthode d'affectation des probabilités possède toutes les propriétés désirables dans le cas d'un seul support. Quels sont les résultats de la méthode lorsque plusieurs supports sont inclus dans le plan publicitaire ? Afin de répondre à cette question, Greene a utilisé son modèle pour estimer la couverture nette et la distribution des occasions de voir de 20 plans publicitaires choisis parmi les 57 plans de la banque de données I.M.S. La table suivante résume les résultats obtenus.

Résumé des performances du modèle des probabilités personnelles pour 20 plans publicitaires

Estimation de la couverture nette

Erreur absolue moyenne sur la couverture nette	=	1 696
Pourcentage moyen d'erreur sur la couverture nette	=	5,9 %
Nombre de sous-estimations de la couverture nette	=	13 sur 20
Nombre d'estimations comprises dans un intervalle de confiance à 95 % autour des données de l'enquête	=	8 sur 20
Erreur type de prévision	=	2 514

Estimation de la distribution des occasions de voir

Chi Deux moyen	=	13,89
Nombre de modes correctement placés	=	20 sur 20
Nombre de pics secondaires correctement placés (2 et 4 expositions)	=	21 sur 40

Commentaires

La méthode des probabilités personnelles est l'une des meilleures méthodes d'affectation des probabilités. Cependant, dans le cas de plusieurs supports, ses performances ne sont pas totalement satisfaisantes. L'erreur absolue moyenne sur la couverture nette est inférieure à celle du modèle de Metheringham, mais elle est supérieure à celle de tous les autres modèles

stochastiques à population hétérogène, comme le montre la table de comparaison de la page suivante. En ce qui concerne le pourcentage moyen d'erreur, la méthode des probabilités personnelles obtient le plus mauvais résultat avec 5,91 % contre 5,2 % à Metheringham. De même, on observe que 8 estimations sur 20 sont à l'extérieur des bornes d'un intervalle de confiance à 95 % autour des données de l'enquête. Néanmoins, la méthode ne paraît pas être systématiquement biaisée puisque l'on observe 7 surestimations pour 13 sous-estimations.

En résumé, il apparaît que l'estimation de la couverture nette par le modèle de simulation n'est pas aussi bonne que les estimations obtenues par les méthodes stochastiques à population hétérogène, Metheringham excepté. Ceci est un résultat important qui vient contredire les affirmations, maintes fois avancées, par les partisans des modèles de simulation quant à la supériorité de leurs modèles.

En ce qui concerne l'estimation de la distribution des occasions de voir, les résultats sont moins touchés. Avec un Chi Deux moyen de 13,89, le modèle des probabilités personnelles s'adjuge la troisième place, loin derrière Modal dont le Chi Deux moyen n'est que de 3,38, mais fort près du modèle de casualité constante dont le Chi Deux moyen est de 10,15. Le modèle de simulation devance donc nettement le modèle de Metheringham et le modèle Multinomial composé en ce qui concerne l'estimation de la distribution des occasions de voir. Néanmoins, la reproduction de la forme de la distribution n'est pas excellente puisque 21 pics secondaires sur 40 sont correctement placés. La meilleure appréciation d'ensemble portée sur la méthode des probabilités personnelles est sans doute celle prononcée par Greene lui-même (34) : "La méthode des probabilités personnelles ne conserve pas la totalité des corrélations positives entre deux supports, mais une amélioration est possible".

La table suivante compare les résultats de six modèles stochastiques à ceux du modèle de simulation de Greene.

Comparaison des Résultats de 6 Modèles Stochastiques et d'un Modèle de Simulation pour 20 Plans Publicitaires

Modèles	Erreur Absolue Moyenne	Pourcentage Moyen d'erreurs	Nombre de sur-estimations	Nombre de sous Estimations	Erreur Type	Dans Intervalle confiance	Chi Deux Moyen
Probabilités Personnelles	1 696	5,91 %	7	13	2 514	8	13,89
Modal II	333	0,75 %	11	9	504	20	3,38
Casualité Constante(1)	378	0,91 %	13	3	495	20	10,15
Multinomial Composé (2)	891	2,23 %	15	5	1 311	17	69,68
Hofmans	1 017	2,54 %	9	11	1 559	16	-
Morgensztern	1 018	2,65 %	8	12	1 575	15	-
Metheringham	2 758	5,20 %	20	0	3 502	10	68,77

(1) - Résultats obtenus pour 16 Plans Publicitaires seulement.

(2) - Modèle développé par l'auteur.

Cette comparaison est la plus complète réalisée à ce jour. Il existe toutefois deux autres études publiées qui se sont attachées à établir la validité des estimations de deux modèles de simulation en les comparant à des données d'enquête ou de panel. Il s'agit du modèle Anglais C.A.M. et du modèle Français S.C.A.L. que nous allons étudier maintenant.

32 - Le modèle de simulation C.A.M.

C.A.M. est l'acronyme de "Computer Assessment of Media Schedules", un modèle de simulation développé par Broadbent, Beale et Hughes pour le "London Press Exchange" (35). Le modèle est construit à partir des données d'audience de télévision fournies par le panel de "Télévision Audience Measurement", Ltd, (TAM), complétées, pour les supports imprimés, par les données d'enquêtes collectées par le "National Readership Survey" (NRS). A l'époque de la construction de CAM, il n'existait pas en Angleterre de données portant sur les habitudes de lecture, mais elles ont été incorporées depuis. Nous ne discuterons pas ici de la méthode utilisée pour combiner ensemble et rendre compatible, dans la mesure du possible, les deux sources de données. Nous étudierons seulement la méthode d'affectation des probabilités utilisée par le modèle et les résultats du test de validation.

C.A.M. fait appel à la méthode actuarielle d'affectation des probabilités en la corrigeant à l'aide d'un ajustement empirique. La journée est divisée en 11 tranches d'écoute télévisée. Pour chaque tranche horaire, le pourcentage de quart d'heures d'exposition mentionné par chaque panéliste constitue la probabilité d''exposition du panéliste pour cette tranche horaire. Chaque panéliste est donc caractérisé par un vecteur de 77 probabilités d'exposition pour une semaine. Ensuite, on calcule une probabilité moyenne pour les 5 jours ouvrables, ce qui réduit le vecteur à 33 probabilités. Celles-ci sont ensuite ajustées. Cet ajustement est nécessaire car, comme nous l'avons montré au chapitre 5, l'affectation actuarielle des probabilités se montre incapable de reproduire les données originales. Le but de l'ajustement est donc de transformer les probabilités actuarielles de façon à ce qu'elles reproduisent les données originales. Nous avons montré au chapitre 5 que ce problème possède une infinité de solutions, ce qui constitue le principal handicap de l'affectation empirique des probabilités. Les constructeurs du modèle CAM ont opté pour un ajustement logistique. Si p est la probabilité actuarielle, la probabilité ajustée p' sera calculée comme suit:

(62) $$p' = pe^x \;/\left[\, 1 + p\,(e^x - 1)\,\right]$$

avec $x = a + by + cy^2 + dy^3$

Les constantes a, b, c, d, sont déterminées tous les 3 mois par la méthode des moindres carrés de façon à ce que la couverture nette simulée soit égale à celle observée par l'intermédiaire du panel pour l'ensemble de la population adulte. Si nous appelons R la couverture nette observée, la variable explicative y de la régression est déterminée comme suit :

(63) $$y = \log\left[\, Q/\,(1 - Q)\,\right] \quad \text{avec } Q = (R - p_a)\;/\,(1 - p_a - p_n)$$

p_a = proportion de panélistes qui déclarent être toujours exposés durant cette tranche horaire.

p_n = proportion de panélistes qui déclarent n'être jamais exposés pendant cette tranche horaire.

Comme on le voit, l'approche utilisée par CAM est totalement empirique. Aucun argument théorique n'est avancé par les auteurs pour justifier le choix de la transformation logistique, si ce n'est qu'elle contraint les probabilités ajustées p' à appartenir à l'intervalle $\left[\, 0,1 \,\right]$.

Afin de valider leur méthode d'affectation des probabilités, les auteurs de CAM ont sélectionné 4 plans publicitaires et déterminé pour chacun d'eux la distribution des occasions de voir directement à partir des livres d'écoute de TAM. Les résultats de ce test sont présentés dans la table suivante.

Test du Modèle de Simulation CAM
à l'aide du panel d'écoute télévisées TAM

Plan Publicitaire	Nombre de Spots	Audience des spots		Distribution des Occasions de Voir 0	1	2	3	4	5	6	7	8	9	10
				32	24	24	15	3	2					
A	5	17 - 42	TAM	33	21	21	17	5	3					
			TAM	14	10	16	14	9	13	12	11	1		
B	11	15 - 53	CAM	14	12	13	15	14	12	10	7	3	1	
			TAM	10	10	16	22	12	18	8	3	1		
C	8	23 - 67	CAM	11	14	16	17	16	13	8	3	2		
			TAM	16	17	10	14	15	12	7	5	1	1	2
D	11	10 - 51	CAM	15	13	15	14	13	11	8	5	3	2	1

Il n'est guère possible de se livrer à des généralisations sur la base d'un échantillon aussi étroit de plans publicitaires. Néanmoins, on peut remarquer que, bien que la couverture soit estimée de façon très satisfaisante, la reproduction des pics secondaires de la distribution n'est pas très bonne. Nous ne savons pas si d'autres tests ont été entrepris sur une plus grande échelle et il n'existe pas, à notre connaissance, d'étude publiée sur le sujet depuis l'expérience de 1966. Toutefois, nous pensons que la méthode ne peut pas conduire à d'aussi bons résultats avec les supports imprimés qui ont une distribution des occasions de voir plus complexe que celle des programmes télévisés. On pourrait aussi considérer que les constructeurs de CAM partagent implicitement cette opinion puisqu'ils ont confiné le test de validation aux seuls programmes télévisés. Ajoutons que CAM simule la distribution des occasions de voir pour les supports imprimés par une simple expansion binomiale des probabilités actuarielles, ce qui constitue, comme nous l'avons montré au chapitre 5, la plus mauvaise solution possible.

Nous allons maintenant présenter les résultats de la deuxième tentative de validation basée sur les données d'un panel, et qui concerne le modèle de simulation Français S.C.A.L.

33 - Le Modèle de Simulation S.C.A.L.

Le modèle de Simulation des Comportements de lecture (S.C.A.L.) a été développé en 1965 par Bertier et Du Jeu (36), en collaboration avec Marc (37). Il ne concerne que la simulation de l'exposition aux supports imprimés. L'affectation des probabilités s'effectue selon un processus en deux étapes qui incorpore une contrainte d'accumulation et une contrainte de duplication.

La nature de ces contraintes et du processus d'affectation des probabilités peut être expliquée par référence au concept de patron de lecture qui décrit le comportement d'exposition d'un individu vis à vis de plusieurs supports au cours des parutions successives de ces supports. Supposons qu'un panéliste ait enregistré sur un journal d'écoute ses expositions à m supports hebdomadaires pour une période de 52 semaines. Son patron de lecture peut être représenté par une table qui comporte 52 colonnes, et m lignes. A l'intersection de la ième ligne avec la jième colonne, nous placerons un 1 si l'individu a été exposé au ième support de la jieme semaine. S'il n'a pas été exposé, nous placerons un zéro.

Patron de lecture

Semaines

Supports	1	2	.....	j	.....	52
⋮						
i				x_{ij}		
⋮						
m						

$X_{ij} = 1$ Exposition

$X_{ij} = 0$ Non exposition

Nous pourrions remplir cette table en tirant des 0 et des 1 au hasard dans une urne contenant des boules de m couleurs différentes, la proportion de boules de la couleur i étant égale à la probabilité d'exposition de l'individu au ième support. Toutefois, en procédant de la sorte, nous admettrions que les audiences de chaque support sont indépendantes, et que les audiences de chaque parution (ou semaine) sont indépendantes. Or, nous savons qu'il n'en est rien. Les duplications, comme les accumulations, ne sont pas aléatoires. Toutes les études d'audience montrent qu'il existe une corrélation positive à la fois entre les supports et entre les parutions d'un même support. Nous devons donc tenir compte de ces deux facteurs en remplissant les cases du patron de lecture. La solution adoptée par le modèle SCAL consiste à contraindre les variances des

lignes et les variances des colonnes à être inférieures aux variances qui résulteraient d'une accumulation et d'une duplication purement aléatoire.

Dans une première étape, le nombre total d'expositions de l'individu est obtenu par tirage au hasard dans une loi normale de moyenne m et de variance σ^2, telle que :

$$m = \pi T \qquad \pi = \Sigma_{i=1}^{m} p_i$$

$$\sigma^2 = T\left[1 - \left(\frac{\pi}{v}\right)^a\right]\left[\Sigma_{i=1}^{m} p_i(1 - p_i)\right]$$

avec : T = Longueur de la période simulée.
π = Somme des probabilités de lecture de l'individu.
p_i = Probabilité de lecture du support i.
v = Nombre de supports pour les lesquels $p_i \neq 0$
a = Paramètre d'ajustement.

Dans une seconde étape, le nombre d'expositions de chaque semaine est obtenu par tirage dans une loi multinomiale dont les probabilités sont définies par p_i / π. Ceci garantit que le nombre d'expositions à chaque support sera globalement proportionnel aux probabilités d'exposition de chaque support.

La validation du modèle a été réalisée en comparant, pour un grand nombre de plans publicitaires, les estimations du modèle avec les observations du panel de lecture, pour une période de 8 semaines. Les résultats de ce test sont présentés dans l'article de Bertier et Du Jeu (36) sous forme graphique et sont extrêmement satisfaisants. La couverture nette est précisément estimée et tous les pics secondaires de la distribution des occasions de voir sont fidèlement reproduits. Cependant, il convient de noter que les paramètres du modèle SCAL ne peuvent pas être estimés à partir des audiences et des duplications. Ils nécessitent l'existence d'un panel. Puisqu'il n'existe pas de panel de lecture pour les supports imprimés aux Etats-Unis, il n'est pas possible de tester le modèle vis à vis des habitudes de lecture américaines.

Malgré ses excellents résultats, le modèle SCAL, qui était utilisé par un grand nombre d'agences de publicité Françaises, est maintenant supplanté par le modèle de Morgensztern que nous avons étudié précédemment. La raison du déclin de SCAL, ne tient pas à la qualité de ses estimations, qui sont excellentes,

mais au fait qu'il ne peut pas être intégré dans un processus d'optimisation. En effet, l'évaluation de la couverture nette et de la distribution des occasions de voir nécessite la simulation des expositions d'un grand échantillon d'individus fictifs. Pour chaque nouveau plan, cette simulation doit être recommencée. Le temps de calcul, ainsi que le coût de l'opération, limite l'utilisateur à l'évaluation de quelques dizaines de plans. Les modèles stochastiques, au contraire, obtiennent le même résultat à l'aide de formules ne faisant appel qu'à des opérations arithmétiques qui n'emploient que quelques fractions de secondes. En conséquence, il devient possible d'utiliser un processus d'optimisation qui évaluera plusieurs milliers de plans publicitaires, pour un coût très raisonnable.

Il existe de nombreux modèles de simulation d'audience actuellement utilisés aux Etats-Unis ou en Europe, mais tous utilisent, soit la méthode des probabilités personnelles, soit la méthode d'affectation empirique et, par conséquent, leurs performances sont probablement comparables à celles observées dans les 3 validations que nous venons d'étudier. On ne peut s'empêcher d'être surpris par le faible nombre de validations auxquelles ont été soumis les modèles de simulation. Les résultats que nous avons présentés montrent que, contrairement aux affirmations de leurs partisans, les modèles de simulation ne sont pas supérieurs aux modèles stochastiques.

Nous avons terminé notre étude des modèles d'exposition complets. Nous sommes maintenant en mesure de définir les divers critères d'évaluation qu'il convient de considérer pour établir un classement des modèles d'exposition.

4 - Evaluation critique des Modèles d'Exposition Complets.

Notre étude suggère qu'un modèle d'exposition satisfaisant doit posséder les caractéristiques suivantes.

a) - Les hypothèses du modèle doivent être énoncées de façon explicite. Elles doivent être cohérentes et parcimonieuses. Elles doivent aussi pouvoir être testées à l'aide des données disponibles. Puisque la détermination du nombre et de la distribution des expositions implique nécessairement que l'on fasse des hypothèses sur le comportement des individus envers les supports et envers les parutions, les critères théoriques qu'il convient de retenir sont les suivants.

Quelle est la nature du phénomène d'exposition ? Est-elle déterministe ou stochastique ? Le comportement d'exposition est-il purement aléatoire ou bien se conf rme-t-il à certaines lois ? Si la seconde alternative est choisie, la loi d'exposition dépend-t-elle des expositions précédentes ou bien est-elle stationnaire dans le temps ? Est-il licite d'agréger les individus, ou convient-il d'introduire explicitement les différences individuelles ? Est-il licite d'agréger les supports ou doit-on maintenir leur propre spécificité et, s'il en est ainsi, doit-on tenir compte de l'interdépendance des supports ? Les paramètres du modèle sont-ils sous-identifiés, sur-identifiés ou exactement identifiés ?

Le modèle reproduit-il les données qui ont servi à estimer ses paramètres ? Les estimations du modèle sont-elles entièrement et uniquement déterminées par les données ou la structure du modèle permet-elle plusieurs estimations également valides ?

b) - Un modèle satisfaisant du point de vue des critères théoriques peut très bien s'avérer inadéquat du point de vue empirique. Un bon modèle doit posséder des conditions d'applications suffisamment souples pour qu'on puisse l'utiliser pour plusieurs catégories de supports. Il doit être robuste afin de ne jamais conduire à des résultats absurdes ou illogiques, telle qu'une couverture nette supérieure à l'unité, des fréquences d'exposition négatives ou encore un déclin de la couverture nette après l'adjonction d'une nouvelle parution ou d'un nouveau support. Les estimations du modèle doivent être aussi précises que possible en toutes circonstances ou, du moins, être toujours comprises dans l'intervalle de confiance des données de l'enquête d'audience. De plus, les estimations du modèle ne doivent être sujettes à aucun biais systématique. Enfin, la distribution des occasions de voir doit être reproduite au mieux, sans effacement ou

disparitions des pics secondaires.

c) - D'excellents résultats, sur le plan empirique, ne suffisent pas. Un modèle satisfaisant ne doit pas nécessiter des calculs trop volumineux afin qu'il puisse être utilisé de façon économique et que l'on puisse, le cas échéant, l'inclure dans un programme d'optimisation. Finalement, le modèle ne doit pas faire appel à d'autres données que celles couramment disponibles, c'est-à-dire les audiences des deux premières parutions de chaque support et les duplications de toutes les paires de supports.

Aucun des modèles étudiés n'est totalement satisfaisant à la fois sur le plan théorique, sur le plan empirique et sur le plan pratique. Néanmoins, certains modèles réalisent un équilibre acceptable.

La table suivante présente une liste de 18 critères qui sont ensuite appliqués au 15 modèles d'exposition complets étudiés dans ce chapitre. Cette seconde table résume, de façon synthétique, les résultats de notre analyse des modèles d'exposition complets.

18 Critères d'Evaluation pour les Modèles d'Exposition Complets

A - Les Critères Théoriques

1 - L'exposition aux supports est-elle considérée comme un phénomène certain ou comme une phénomène aléatoire ?
2 - Les probabilités d'exposition sont-elles stationnaires d'une parution à une autre ?
3 - Les probabilités d'exposition sont-elles stationnaires d'un individu à un autre ?
4 - Les audiences des supports sont-elles considérées indépendantes ?
5 - Le phénomène d'accumulation se produit-il au pur hasard ?
6 - Les individus sont-ils considérés comme homogènes ?
7 - Le modèle permet-il l'estimation de la distribution des occasions de voir ?
8 - Les paramètres du modèle sont-il exactement identifiés ?
9 - Le modèle recourt-il à des approximations ?
10 - Le modèle reproduit-il les données qui ont servi à estimer ses paramètres ?
11 - Les estimations produites sont-elles uniques ?

B - Les Critères Empiriques

12 - Les conditions d'application sont-elles difficiles à remplir ?
13 - Le modèle est-il robuste ?
14 - Les estimations du modèle sont-elles soumises à un biais systématique ?
15 - L'erreur de prédiction est-elle inférieure à l'erreur d'échantillonnage ?
16 - Le modèle reproduit-il fidèlement la distribution des occasions de voir ?

C - Les Critères Pratiques

17 - Le modèle nécessite-t-il des calculs volumineux ?
18 - Le modèle nécessite-til des données d'audience non habituellement disponibles ?

Evaluation de 15 Modèles d'Exposition Complets

Critères d'évaluation

	THEORIQUES											EMPIRIQUES					PRATIQUES	
Modèles Stochastiques	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Binomial	A	0	0	0	0	0	0	0	N	0	0	N	N	SU	N	N	N	N
Binomial multivarié	A	0	0	0	0	0	0	0	N	0	0	N	N	SU	N	N	N	N
Multinomial *	A	0	0	N	0	0	0	0	N	0	0	N	0	SU	N	N	0	N
Binomial Markovien *	A	N	0	N	N	0	0	0	N	0	0	N	N	SU	N	N	0	N
Metheringham	A	0	0	N	N	N	0	0	N	N	0	N	N	SU	N	N	N	N
Bêta Matriciel	A	0	0	N	N	N	0	0	N	0	0	0	0	N	0	0	0	0
Multinomial composé *	A	0	0	N	N	N	0	0	N	0	0	N	0	SU	N	0	0	N
Modal	A	0	0	N	N	N	0	N	0	0	0	N	N	N	0	0	0	N
Hofmans	C	0	0	N	N	N	N	0	0	0	0	0	N	SO	N	NA	N	0
Morgenstern	A	0	0	N	N	N	0	0	0	0	0	0	0	SO	N	0	N	0
Casualité constante	A	0	0	N	N	N	0	N	0	0	0	N	?	SU	0	0	0	N
Metheringham-Multinomial composé *	A	0	0	N	N	N	0	N	0	N	0	N	N	SU	N	0	0	N
Modèles de Simulation																		
Probabilités Personnelles	A	0	0	N	N	N	0	0	N	0	0	N	0	SO	N	N	0	N
C.A.M.	A	0	0	N	N	N	0	N	0	0	N	N	0	?	?	N	0	N
SCAL	A	0	0	N	N	N	0	N	0	0	0	N	0	?	?	?	0	0

* Modèles développés par l'auteur

Définition des Symboles

0 = Oui
N = Non
? = Inconnu

C = Certain
A = Aléatoire
NA = ne s'applique pas

SU = Surestimation
SO = Sous-estimation

Le classement des modèles d'exposition est une tâche difficile. Les résultats d'un tel classement dépendent beaucoup de l'importance relative attribuée à chaque critère. Par exemple, si l'accent est mis sur les critères théoriques, on tendra à rejeter les modèles qui recourent à des approximations et qui sont sous-identifiés. Par contre, si l'accent est mis sur les performances empiriques, les modèles Modal et de casualité constante seront très certainement préférés. Si la stationnarité des probabilités est considérée comme étant une hypothèse trop restrictive, aucun des modèles stochastiques n'apparaîtra satisfaisant. Par contre, si l'on attache un grand prix à la rapidité des calculs, alors aucun modèle de simulation n'apparaîtra comme satisfaisant. Nous pourrions continuer cette énumération, mais nous pensons qu'il convient de laisser à chaque utilisateur le privilège de choisir ses propres priorités et d'établir la pondération des critères qui en résultent. Néanmoins, nous allons tout de même indiquer les modèles qui sont clairement dominés pour l'ensemble des critères et qui, par conséquent, ne devraient pas être utilisés.

La liste des modèles dominés comprend les quatre modèles à population homogène qui échouent, à la fois sur les critères théoriques et sur les critères empiriques. Le modèle de Metheringham est également dominé par tous les autres modèles à population hétérogène. Il peut être remplacé avantageusement par le modèle multinomial composé si un modèle à distribution exacte est désiré, ou par le modèle Modal ou le modèle de Casualité constante si l'on souhaite obtenir une précision maximum.

Finalement, parmi les modèles de simulation, les méthodes actuarielles et empiriques d'affectation des probabilités devraient être abandonnées car elles conduisent à des estimations systématiquement biaisées. Quand les données disponibles se limitent aux audiences des 2 premières parutions et aux duplications par paire, la méthode des probabilités personnelles constitue probablement le meilleur choix. Toutefois, quand des données de panel sont disponibles, l'affectation des probabilités selon la méthode SCAL permet un meilleur ajustement.

Nous avons achevé notre analyse des modèles d'exposition complets. Nous conclurons notre étude en résumant brièvement les diverses contributions que nous avons apportées à une meilleure connaissance des phénomènes d'exposition aux supports publicitaires et en indiquant les directions probables que prendront les recherches futures.

5 - Conclusion et Directions pour des recherches futures.

Dans ce chapitre, nous avons montré que la relation liant l'accumulation d'audiences à la distribution des occasions de voir peut être étendue à des plans publicitaires comportant plusieurs supports. Nous avons établi une relation récursive générale qui permet de dériver les matrices d'exposition réduites à partir de la matrice d'exposition complète. Nous avons illustré la correspondance qui existe entre la distribution jointe des occasions de voir et la matrice d'exposition. Nous avons décrit les caractéristiques de la banque de données utilisée pour tester les 15 modèles d'exposition complets et souligné ses limitations.

Nous avons présenté ensuite un schéma général de classification des modèles d'exposition complets. Cette classification sépare les modèles en deux grandes catégories : ceux qui assument que les individus sont homogènes et ceux qui s'efforcent de tenir compte des différences individuelles. Quatre modèles appartenant à la première catégorie ont été étudiés en détail, leurs hypothèses ont été examinées et leurs principales équations présentées. Les problèmes posés par l'estimation de leurs paramètres ont été discutés et des exemples numériques ont été fournis. Les résultats respectifs de ces 4 modèles, en terme de couverture nette et de distribution des occasions de voir ont été déterminés à l'aide des 57 plans publicitaires de la banque de données. Pour chaque modèle, un résumé des performances a été présenté sous forme standardisée afin de faciliter les comparaisons.

Ensuite, huit modèles à population hétérogène ont été étudiés : certains très connus comme le modèle de Metheringham, d'autres moins connus, comme le modèle de casualité constante, d'autres enfin entièrement nouveaux comme notre modèle Multinomial composé. Les trois premiers d'entre eux sont des modèles à distribution exacte, les cinq suivants sont des modèles d'approximation. Pour chacun d'eux, nous avons présenté leurs équations et montré, à l'aide d'exemples numériques, comment leurs paramètres peuvent être estimés. En ce qui concerne les deux modèles couverts par le secret commercial, nous avons complété le peu d'informations publiées à leur sujet par nos propres déductions. De nouveau, les résultats de chacun des modèles ont été présentés et analysés sur la base d'un grand nombre de critères.

Nous avons ensuite étudié les 3 études de validation réalisées dans le domaine des modèles de simulation. La validation la plus approfondie a été

conduite sur notre banque de données permettant ainsi la première comparaison à armes égales entre modèles stochastiques et modèles de simulation. L'issue de cette comparaison défavorable au modèle de simulation, constitue en soi une contribution fort importante contredisant les affirmations verbales de supériorité des partisans des modèles de simulation.

Notre principale contribution réside dans le rapprochement de 15 modèles d'exposition complets dont quatre entièrement nouveaux et dans la réalisation de la première comparaison exhaustive de leurs résultats à l'aide d'une banque de données unifiées. Cette comparaison permet, pour la première fois, une évaluation objective des mérites de chacun des modèles.

Notre seconde contribution réside dans la construction d'un nouveau modèle dont les résultats sont supérieurs à ceux du modèle de Metheringham, sans pour autant nécessiter le recours à des données supplémentaires. Cette construction a été rendue possible par un effort préalable de classification des hypothèses, implicites ou explicites des modèles préexistants. Il ne sera désormais plus possible de prétendre que l'exposition aux supports est un phénomène purement aléatoire. L'existence de la fidélité des audiences d'une parution à une autre et l'interdépendance des audiences d'un support à un autre ont été démontrées par le succès des modèles incorporant ces deux phénomènes et par l'échec des modèles qui n'en tiennent pas compte.

Notre troisième contribution tient à la production d'une grille d'évaluation multicritères permettant aux utilisateurs de choisir le modèle qui convient le mieux à leurs objectifs. Cette grille d'évaluation nous a aussi permis d'établir une liste de modèles dominés dont l'utilisation devrait être proscrite.

Finalement, notre étude a révélé une limitation commune à la plupart des modèles disponibles : l'hypothèse de stationnarité des probabilités de lecture. Notre première tentative de construction d'un modèle non stationnaire, le modèle Binomial Markovien, ouvre probablement la voie à la construction de modèles plus complexes capables de tenir compte, à la fois, de l'hétérogénéité des individus et de la variabilité dans le temps de leurs probabilités d'exposition. La tendance à l'extension aux media imprimés des techniques de panel, jusqu'ici réservées aux media radiodiffusés, est certainement appelée à se développer, non seulement en Europe, où elle a pris naissance, mais aussi aux Etats-Unis.

Les modèles d'exposition entrent dans leur deuxième décade d'existence. Les dix dernières années ont été le témoin de progrès rapides, spécialement marqués dans le domaine empirique par l'accroissement important de la précision des estimations. Les modèles ont évolué. Les simples formules du début ont laissé la place à des modèles stochastiques complexes permettant d'intégrer les différences entre supports comme les différences entre individus. Par contre, l'effet des expositions passées, des fluctuations tendancielles et saisonnières d'audience n'est toujours pas pris en considération. Il est certain que ces phénomènes feront l'objet d'études poussées au cours des dix prochaines années.

N O T E S

(1) - La coopération de M. Harold ISRAEL, Vice Président de la Société W.R. Simmons et Associés, et de M. Léon H. LIEBMAN, Président de la Société Interactive Market Systems, qui ont fourni respectivement les données d'enquête et leur tabulation, a rendu possible cette étude. Nous leur exprimons ici toute notre reconnaissance.

(2) - POLITZ Alfred Research Inc., "Life Study of Consumer Expenditures", 1959, New-York.

(3) - POLITZ Alfred Media Studies, "Main Report, 1964 Magazine Study, the audiences of 11 magazines", Février 1965, New-York.

(4) - LEE A.M., BURKART A.J., "Some optimization problems in advertising media planning", Operational Research Quaterly, vol. 11, n° 3, Septembre 1960.

(5) - KENDALL et STUART, "Advanced Theory of Statistics", vol. 1, Griffin, Londres, 1958.

(6) - STEYN H.S., "On the univariate series $F(t) = F(a, b_1, b_2, \ldots, b_k, c, t, t^2, \ldots, t^k)$ and its applications to probability theory," Nederland Akademia Wetensch., Proceding Series A, 58, pp. 588-595.

(7) - KLOTZ J., "Statistical Inference in Bernoulli Trials with Dependence", The Annals of Statistics, 1973, vol. 1, pp. 373-379.

(8) - ISHII G. et HAYAKAWA R, "On the Compound Binomial Distribution", Annals of the Statistic Mathematics, vol. 12., pp. 69-80, Tokyo.

(9) - CHATFIELD C. et GOODHART G., "The Bêta Binomial Model for Consumer Purchasing Behaviour", Applied Statistics, vol. 19, Part 3, pp. 240-250, 1970.

(10)- METHERINGHAM R.A., "Measuring the Net Cumulative Coverage of a Print Campaign", Journal of Advertising Research", Décembre 1969, pp. 23-28.

(11) - GREENE J.D., "Personal Media Probabilities", Journal of Advertising Research, vol. 10, N° 5, Octobre 1970, pp. 12-18.

(12) - SCHREIBER R.J., "The Metheringham Method for Media Mix : An Evaluation", Journal of Advertising Research, vol. 9, n° 2, pp. 54-56, 1969.

(13) - COMET est l'acronyme de : Computerized Metheringham Technique.

(14) - COSMIC, A comprehensive Press Planning System, Foote, Cone et Belding, Londres, Septembre 1970.

(15) - GREENE J.D., STOCK J.S., Advertising Reach and Frequency in Magazines, Marketmath Inc. and Reader's Digest Association, New-York, 1967.

(16) - Simmons W.R. & Associates Research, Inc., Mass Markets and the Media Reaching Them, 1966 Magazine Reach and Frequency Report, New-York, 1966.

(17) - DEMING M.E. et STEPHAN F.F., "On a least square adjustement of a sampled frequency table when the expected marginals are known", Annals of Mathematical Statistics, vol. 11, pp. 427-444, 1940.

(18) - Le Modèle Multinomial composé pourrait aussi bien recevoir le nom de modèle Dirichlet Multinomial.

(19) - MAULDON J.G., "A generalization of the Beta Distribution", Annals of Mathematical Statistics, vol. 30, 1959, pp. 509-520.

(20) - MOSSIMANN J.E., "On the compound multinomial distribution, the multivariate distribution, and correlations among proportions", Biometrika, vol. 49, 1962, pp. 65-82.

(21) - JOHNSON N.L., KOTZ S., Discrete Distributions, Houghton Mifflin, Boston, 1969, Chapter 11.

(22) - JOHNSON N.L., KOTZ S., Continuous Multivariate Distributions, J. Wiley & Sons, 1972, Chapter 40.

(23) - Interactive Market Systems, Inc, 19 West 44 th Street, New-York.

(24) - LIEBMAN L.H., "Modal reach and frequency : An improved estimation method", Proceedings of the 17 th Annual A.R.F. Conference, November 1971.

(25) - Metheringham R.A., "Models, Media measurement and the Metheringham formula", Proceedings of the 17 th Annual A.R.F. Conference, Novembre 1971.

(26) - Interactive Market Systems, A Comparaison of available reach and frequency procedures : Final report, I.M.S., February 1973.

(27) - HOFMANS P., "Measuring the cumulative net coverage of any combination of media", Journal of Marketing Research, Vol. 3, August 1966, pp. 269-278.

(28) - Morgenztern A., "Fréquentation de la presse, généralisation, effet publicitaire", Etude I.R.E.P., n° 19, Avril 1970.

(29) - Fry C., "The concept of casualness", Feedback, n° 2, April 1973.

(30) - Christopher Fry, PTY Limited, Statistical Applications, P8/287, Military Road, Gremorne 2090.

(31) - Gensch D.H., "A computer simulation model for selecting advertising schedules", Journal of Marketing Research, May 1969, pp. 203-214.

(32) - AAKER D.A., "ADMOD : An advertising decision model", Journal of Marketing Research, February 1975, pp. 37-45.

(33) - FRIEDMAN L., "Constructing a media simulation model", Journal of Advertising Research, August 1970, pp. 33-39.

(34) - GREENE J.D., "Personal Media Probabilities", Journal of Advertising Reserach, October 1970, pp. 12-18.

(35) - BEALE E.M.L., HUGHES P.A.B., BROADBENT S.R., "A computer Assessment of Media Schedules", Operational research Quaterly, vol. 17, December 1966, pp. 381-411.

(36) - BERTHIER, DU JEU P., "Simulation des comportements aléatoires de lecture : le modèle SCAL", Metra, vol. VI, n° 4, 1967, pp. 647-659.

(37) - MARC M., "Combining simulation and panel date to obtain reach and frequency", Journal of Advertising Research, vol. 8, N° 2, June 1968, pp. 11-16.

APPENDICES

-:-:-:-:-:-:-:-:-:-

BANQUE DE DONNEES

ET

DISTRIBUTION DES OCCASIONS DE VOIR

POUR NEUF MODELES D'EXPOSITION COMPLETS

A = Liste des codes de magazine

B = Liste des plans publicitaires

C = Classification des magazines

D = Classification des plans publicitaires

E = Distributions observées et estimées des occasions de voir

F = Bibliographie

A P P E N D I C E S

-:-:-:-:-:-:-:-:-:-

BANQUE DE DONNEES

ET

DISTRIBUTION DES OCCASIONS DE VOIR

POUR NEUF MODELES D'EXPOSITION COMPLETS

A = Liste des codes de magazine

B = Liste des plans publicitaires

C = Classification des magazines

D = Classification des plans publicitaires

E = Distributions observées et estimées des occasions de voir

F = Bibliographie

APPENDICE A

LISTE DES CODES DE MAGAZINE

AMH	American Home
ARG	Argosy
BAR	Barron's
BET	Better Homes & Gardens
BWK	Business Week
COS	Cosmopolitan
DUN	Dun's Review
ESQ	Esquire
FAM	Family Circle
FWK	Family Weekly
FIE	Field & Stream
FBS	Forbes
FOR	Fortune
GLA	Glamour
GOO	Good Housekeeping
GUN	Guns & Ammo
HAT	Harper's/Atlantic
HOT	Hot Rod
HOU	House & Garden
HSB	House Beautiful
LAD	Ladies' Home Journal
LIF	Life
MCC	McCall's
MEC	Mechanix Illustrated
NAT	Nation's Business
NGE	National Geographic
NOB	National Observer
NWK	Newsweek
NEW	New Yorker
OUT	Outdoor Life
PAR	Parade
PLA	Playboy
POP	Popular Mechanics
PSM	Popular Science
PSY	Psychology Today
REA	Reader's Digest
RED	Redbook
SAT	Saturday Review
SPO	Sport
SAF	Sports Afield
SIM	Sports Illustrated
TIM	Time
TST	True Story
TVG	TV Guide
USN	U.S. News & World Report
WSJ	Wall Street Journal
WOM	Woman's Day

APPENDICE B

LISTE DES PLANS PUBLICITAIRES

1. LIF-REA-TVG-TIM
2. LIF-REA-TVG-BET
3. PAR-REA-TVG-BET
4. FWK-GOO-NGE-PLA
5. NWK-SIM-NGE-MCC
6. NWK-SIM-FAM-RED
7. AMH-ESQ-MEC-SPO
8. FIE-ESQ-POP-SPO
9. HOU-PSM-OUT-AMH
10. HAT-NEW-FOR-SAT
11. BWK-COS-FOR-SAT
12. GLA-HSB-NAT-PSY
13. FOR-AMH-NGE-REA
14. HAT-OUT-PLA-TVG
15. HSB-ESQ-NWK-REA
16. ARG-FOR-SIM-WSJ
17. GUN-HSB-NAT-SIM
18. BWK-FOR-NAT-WSJ
19. COS-HAT-NOB-SAT
20. LIF-NWK-NEW-SAT
21. LIF-NOB-DUN-PSY
22. AMH-BET-ESQ-LAD
23. MCC-PLA-RED-TIM
24. TIM-FBS-AMH
25. FWK-PAR-REA-SPO
26. FWK-PAR-NGE-TVG
27. REA-PAR-NGE-TVG
28. PAR-FAM-LIF-FOR
29. REA-GOO-TIM-HSB
30. NGE-RED-NWK-SIM
31. SAT-NEW-HAT
32. BAR-FOR-WSJ-BWK
33. ARG-OUT-SAF-FIE
34. BET-FAM-LAD
35. BET-FAM-MCC
36. LIF-PLA-NEW
37. ESQ-SPO-POP
38. PAR-TVG-REA
39. AMH-FWK-HOT-TST
40. TIM-FAM-LIF
41. WOM-FAM-SIM-NEW
42. WOM-FAM-NEW
43. FAM-NEW-MCC
44. HAT-SAT-LIF
45. HAT-SAT-LIF-TVG
46. LIF-REA-SIM-TVG
47. FAM-MCC-WOM-LAD
48. TIM-USN-NWK-SAT
49. LIF-REA-SIM-TVG
50. FAM-MCC-WOM-LAD
51. TIM-USN-NWK-SAT
52. LIF-REA-SIM-TVG
53. FAM-MCC-WOM-LAD
54. TIM-USN-NWK-SAT
55. LIF-REA-SIM-TVG
56. FAM-MCC-WOM-LAD
57. TIM-USN-NWK-SAT

APPENDICE C

CLASSIFICATION DES MAGAZINES SELON LEUR AUDIENCE

très faible audience $R_1 < 3,75\%$	faible audience $3,75\% < R_1 < 7,5\%$	audience moyenne $7,5\% < R_1 < 15\%$	audience forte $R_1 > 15\%$
ARG	AMH	FAM	LIF
BAR	ESQ	FWK	PAR
BWK	FIE	GOO	REA
COS	HOU	LAD	TVG
DUN	MEC	MCC	TIM
FBS	OUT	NGE	BET
FOR	POP	NWK	
GLA	PSM	PLA	
GUN	SPO	RED	
HAT	SAF	SIM	
HOT	TST	WOM	
HSB	USN		
NAT			
NOB			
NEW			
PSY			
SAT			
WSJ			

CLASSIFICATION DES MAGAZINES SELON LEUR TAUX D'ACCUMULATION

très faible	faible	moyen	fort
t < 32,5%	32,5% < t < 40%	40% < t < 47,5	t > 47,5%
FWK	AMH	COS	ARG
PAR	BET	DUN	BAR
REA	ESQ	GLA	BWK
SPO	FAM	GOO	FIE
NGE	FBS	HAT	FOR
TVG	LAD	HOT	GUN
	MCC	HOU	HSB
	PLA	LIF	NAT
	RED	MEC	OUT
	SAF	NOB	SIM
	TIM	NWK	WSJ
	USN	NEW	
		POP	
		PSM	
		PSY	
		SAT	
		TST	
		WOM	

CLASSIFICATION DES MAGAZINES SELON LEUR DUPLICATION

faible	moyenne	forte
PAR-TVG	BET-FAM	SAT-NEW
PAR-REA	BET-LAD	SAT-HAT
REA-TVG	FAM-LAD	NEW-HAT
AMH-FWK	BET-MCC	BAR-FOR
AMH-HOT	FAM-MCC	BAR-BWK
AMH-TST	LIF-PLA	BAR-WSJ
FWK-HOT	LIF-NEW	FOR-BWK
FWK-TST	PLA-NEW	FOR-WSJ
HOT-TST	ESQ-SPO	BWK-WSJ
TIM-FAM	ESQ-POP	ARG-OUT
TIM-LIF	POP-SPO	ARG-SAF
FAM-LIF		ARG-FIE
		OUT-SAF
		OUT-FIE
		SAF-FIE

CLASSIFICATION DES POPULATIONS CIBLES

Très petite, moins 10,000

Men, Professional - Managerial, Co. Size 500 and Over

Petite, 10,001 - 25,000

Women, Attended/Graduated College

Moyenne, 25,001 - 500,00

Women, HH Income $ 10,000 and Over

Grande, plus de 50,000

Total Women

APPENDICE D

CLASSIFICATION DES PLANS PUBLICITAIRES

Plans	
1 - 3	Large Audience
4 - 6	Audience Moyenne
7 - 9	Audience faible
10 - 12	Audience très faible
13 - 15	Audiences mélangées
16 - 18	Taux d'Accumulation élevé
19 - 21	Taux d'Accumulation moyen
22 - 24	Taux d'Accumulation faible
25 - 27	Taux d'Accumulation très faible
28 - 30	Taux d'Accumulation mélangés
31 - 33	Duplications élevées
34 - 37	Duplications moyennes
38 - 40	Duplications faibles
41 - 45	Duplications mélangées
46 - 48	Large population cible
49 - 51	Population cible moyenne
52 - 54	Population cible faible
55 - 57	Population cible très faible

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 1

	1 ISSUE	2 ISSUES
LIF	33301	47669
REA	42616	52989
TVG	35821	46874
TIM	20355	27400

NET AND DUPLICATED AUDIENCES

	LIF	REA	TVG	TIM
LIF	33301	14136	11584	9532
REA	61781	42616	14093	9757
TVG	57538	64344	35821	7345
TIM	44124	53214	48831	20355

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	34815	26621	28581	16698	14154	6200	4892	1590	776	95511
BINOMIAL	12583	34164	40582	27546	11686	3173	538	52	2	117743
MULTI-BIN	12249	34132	41040	27781	11565	3027	486	44	2	118077
MULTINOMIA	17736	34422	34182	24009	12879	5156	1555	345	42	112590
MARKOV-BIN	18840	33234	33806	24174	12968	5301	1618	343	41	111486
METHERINGH	32280	29337	23673	17752	12376	7887	4424	2011	585	98046
COMP-MULTI	34934	15279	39699	11221	19658	2351	5936	138	1110	95392
MODAL	35135	21996	30117	15138	14176	6377	4780	1587	1020	95191
CASUALNESS	34827	20791	27954	19561	15033	6823	3466	1186	680	95499
METH-C.M.	34453	22197	30992	14369	15852	5158	5295	1084	925	95873

SCHEDULE 2

	1 ISSUE	2 ISSUES
LIF	33301	47669
REA	42616	52989
TVG	35821	46874
BET	22724	30249

NET AND DUPLICATED AUDIENCES

	LIF	REA	TVG	BET
LIF	33301	14136	11584	7971
REA	61781	42616	14093	10751
TVG	57538	64344	35821	7697
BET	48054	54589	50848	22724

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	33682	21068	29323	18092	14794	6733	5011	967	658	96644
BINOMIAL	11983	33322	40538	28181	12244	3405	592	59	3	118343
MULTI-BIN	11727	33290	40891	28372	12154	3290	549	52	2	118599
MULTINOMIA	16110	34126	35146	24694	13137	5210	1532	331	40	114216
MARKOV-BIN	17592	32775	34308	24873	13370	5417	1621	332	37	112734
METHERINGH	30078	29593	24567	18536	12815	8001	4343	1883	512	100248
COMP-MULTI	33191	15024	40818	11550	20329	2361	5906	125	1023	97135
MODAL	32969	22479	30447	15807	15002	6597	4829	1400	795	97357
CASUALNESS	32978	20958	28717	20040	15424	6940	3489	1140	636	97348
METH-C.M.	32444	22208	32027	14915	16414	5220	5245	1013	840	97882

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 3

	1 ISSUE	2 ISSUES
PAR	32417	39545
REA	42613	52989
TVG	35821	46874
BET	22724	30249

NET AND DUPLICATED AUDIENCES

	PAR	REA	TVG	BET
PAR	32417	13000	11253	7032
REA	62030	42613	14093	10751
TVG	56985	64341	35821	7697
BET	48109	54586	50848	22724

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	35779	18628	30072	15741	17492	6535	4530	1057	492	94547
BINOMIAL	12205	33637	40558	27945	12034	3317	571	56	2	118121
MULTI-BIN	11943	33610	40918	28132	11940	3202	529	49	2	118383
MULTINOMIA	15717	34576	35792	24729	12835	4949	1403	290	34	114609
MARKOV-BIN	17882	32997	34280	24695	13201	5324	1587	325	36	112444
METHERINGH	30464	29667	24495	18418	12703	7918	4294	1861	506	99862
COMP-MULTI	35243	12457	42531	9515	21360	1833	6262	87	1037	95083
MODAL	34213	20028	32602	14535	16038	5959	5038	1162	751	96113
CASUALNESS	34649	19139	29941	18399	16136	6505	3742	1087	724	95677
METH-C.M.	35274	20594	31433	13480	16623	4965	5845	1004	1107	95052

SCHEDULE 4

	1 ISSUE	2 ISSUES
FWK	16783	20834
GOO	15845	22532
NGE	15842	20841
PLA	16377	22561

NET AND DUPLICATED AUDIENCES

	FWK	GOO	NGE	PLA
FWK	16783	2360	2415	5416
GOO	30268	15845	2737	1779
NGE	30210	28950	15842	3425
PLA	27744	30443	28794	16377

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	65783	24547	24464	8193	5746	949	613	26	7	64543
BINOMIAL	45030	51178	25447	7230	1284	146	10	0	0	85296
MULTI-BIN	45028	51180	25448	7230	1284	146	10	0	0	85298
MULTINOMIA	50293	45234	23290	8721	2281	438	60	4	0	80033
MARKOV-BIN	52111	43245	22759	8810	2647	624	113	15	1	78215
METHERINGH	65230	30977	16854	9126	4694	2206	897	286	55	65096
COMP-MULTI	68750	17954	31302	3387	7349	181	1275	3	126	61576
MODAL	66011	23227	26544	1515	5148	1183	586	81	30	64315
CASUALNESS	66246	23159	25777	8386	4949	1129	461	114	100	64080
METH-C.M.	70009	22418	22447	6103	6390	1216	1421	145	177	60317

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 5

	1 ISSUE	2 ISSUES
NWK	13690	20029
SIM	10987	16409
NGE	15842	20841
MCC	17653	23654

NET AND DUPLICATED AUDIENCES

	NWK	SIM	NGE	MCC
NWK	13690	943	1144	581
SIM	23734	10987	2017	1274
NGE	28388	24812	15842	2935
MCC	30762	27366	30560	17653

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	73279	21800	21414	7081	4325	1598	653	149	28	57047
BINOMIAL	50576	50821	22342	5613	881	89	6	0	0	79750
MULTI-BIN	50483	50936	22368	5583	866	86	5	0	0	79843
MULTINOMIA	49496	52432	22290	5205	806	91	7	0	0	80830
MARKOV-BIN	53801	46859	21467	6527	1420	225	25	2	0	76525
METHERINGH	61308	38534	18819	7821	2784	829	196	33	3	69018
COMP-MULTI	67731	23069	31547	2715	4765	54	428	0	17	62595
MODAL	73352	22124	21180	6791	4280	1546	807	178	68	56974
CASUALNESS	73084	22163	20741	8007	4262	1236	505	184	139	57242
METH-C.M.	65951	29366	25166	5110	3925	437	342	16	12	64375

SCHEDULE 6

	1 ISSUE	2 ISSUES
NWK	13690	20029
SIM	10987	16409
FAM	14516	20293
RED	10473	14412

NET AND DUPLICATED AUDIENCES

	NWK	SIM	FAM	RED
NWK	13690	943	478	410
SIM	23734	10987	1136	869
FAM	27728	24367	14516	3364
RED	23753	20591	21625	10473

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	78488	22380	18474	6025	3636	732	481	47	64	51838
BINOMIAL	58502	49285	18165	3826	504	42	2	0	0	71824
MULTI-BIN	58452	49353	18170	3810	498	41	2	0	0	71874
MULTINOMIA	58680	49044	18058	3948	546	46	2	0	0	71646
MARKOV-BIN	61854	44792	17775	4803	948	139	15	1	0	68472
METHERINGH	69417	35762	15888	6269	2171	639	151	26	2	60909
COMP-MULTI	73484	23754	27698	1867	3276	22	220	0	6	56842
MODAL	78459	22936	17600	6077	3529	1070	472	71	22	51867
CASUALNESS	78385	22465	18353	6595	3116	860	332	126	90	51941
METH-C.M.	72443	28937	21690	3984	2742	326	188	13	4	57883

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 7

	1 ISSUE	2 ISSUES	NET AND DUPLICATED AUDIENCES AMH	ESQ	MEC	SPO
AMH	9320	12566	9320	706	466	213
ESQ	7306	10058	15920	7306	828	782
MEC	5036	7142	13890	11514	5036	540
SPO	5916	7541	15023	12440	10412	5916

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	98681	14014	13835	2041	1501	199	49	6	2	31645
BINOMIAL	84370	37701	7371	823	57	3	0	0	0	45956
MULTI-BIN	84312	37798	7348	809	55	2	0	0	0	46014
MULTINOMIA	86258	34589	8054	1252	155	14	1	0	0	44068
MARKOV-BIN	88379	31321	8462	1801	313	45	5	0	0	41947
METHERINGH	95858	21424	8022	3196	1231	432	129	30	4	34468
COMP-MULTI	98532	13116	16253	487	1781	5	145	0	6	31794
MODAL	98651	14074	13812	2143	1327	209	99	8	3	31675
CASUALNESS	98730	14082	13610	2354	1213	185	78	32	39	31596
METH-C.M.	98836	16090	11511	1831	1638	214	181	14	9	31490

SCHEDULE 8

	1 ISSUE	2 ISSUES	NET AND DUPLICATED AUDIENCES FIE	ESQ	POP	SPO
FIE	6987	10379	6987	1003	1083	751
ESQ	7306	10058	13290	7306	566	782
POP	5565	8098	11469	12305	5565	456
SPO	5916	7541	12152	12440	11025	5916

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	101483	13413	10984	2367	1744	299	117	8	2	28843
BINOMIAL	86868	36146	6580	685	45	2	0	0	0	43458
MULTI-BIN	86857	36166	6575	682	44	2	0	0	0	43469
MULTINOMIA	90681	29887	7908	1586	234	27	2	0	0	39645
MARKOV-BIN	91608	28587	7918	1803	346	56	8	1	0	38718
METHERINGH	99625	18314	7155	3107	1338	538	189	52	9	30701
COMP-MULTI	100885	12961	13613	685	1905	13	245	0	19	29441
MODAL	101422	13645	10675	2559	1477	355	161	23	8	28904
CASUALNESS	101479	13409	10933	2689	1277	295	127	58	54	28847
METH-C.M.	102765	13443	9730	1873	1861	271	325	25	33	27561

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 9

	1 ISSUE	2 ISSUES
HOU	5389	7635
PSM	5180	7266
OUT	6372	9626
AMH	9320	12566

NET AND DUPLICATED AUDIENCES

	HOU	PSM	OUT	AMH
HOU	5389	346	398	1213
PSM	10223	5180	793	544
OUT	11363	10759	6372	424
AMH	13496	13956	15268	9320

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	99526	14720	12278	2399	1125	133	139	7	1	30800
BINOMIAL	86188	36577	6791	721	48	2	0	0	0	44138
MULTI-BIN	86126	36681	6765	706	46	2	0	0	0	44200
MULTINOMIA	88662	32514	7593	1383	159	14	1	0	0	41664
MARKOV-BIN	90120	30261	7955	1661	285	40	5	0	0	40206
METHERINGH	97401	20562	7612	3022	1164	410	123	28	4	32925
COMP-MULTI	99012	14617	14322	586	1622	7	153	0	8	31314
MODAL	99473	14132	11809	2346	1221	238	93	11	3	30853
CASUALNESS	99498	15052	11892	2464	1063	200	80	37	36	30828
METH-C.M.	99647	16506	10568	1772	1449	203	159	14	8	30679

SCHEDULE 10

	1 ISSUE	2 ISSUES
HAT	2914	4197
NEW	3294	4712
FOR	2307	3411
SAT	2105	3038

NET AND DUPLICATED AUDIENCES

	HAT	NEW	FOR	SAT
HAT	2914	548	369	651
NEW	5660	3294	380	569
FOR	4852	5221	2307	283
SAT	4368	4830	4129	2105

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	118799	6030	3311	971	720	240	197	56	2	11527
BINOMIAL	110540	18390	1339	56	1	0	0	0	0	19786
MULTI-BIN	110534	18401	1334	55	1	0	0	0	0	19792
MULTINOMIA	114395	11625	3441	743	109	11	2	0	0	15931
MARKOV-BIN	114026	12429	3015	680	142	27	5	1	0	16300
METHERINGH	118564	6703	2626	1267	640	316	143	54	13	11762
COMP-MULTI	118455	5518	4908	356	872	6	184	0	27	11871
MODAL	119090	5493	3426	1139	692	277	148	41	18	11236
CASUALNESS	118802	5543	3835	1200	578	193	93	47	31	11524
METH-C.M.	119929	5275	3297	754	711	145	166	24	24	10397

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 11

	1 ISSUE	2 ISSUES
BWK	4349	6552
COS	4814	6841
FOR	2307	3411
SAT	2105	3038

NET AND DUPLICATED AUDIENCES

	BWK	COS	FOR	SAT
BWK	4349	253	507	437
COS	8910	4814	179	207
FOR	6149	6942	2307	283
SAT	6017	6712	4129	2105

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	113803	9035	5510	1077	698	154	46	2	2	16523
BINOMIAL	105526	22571	2112	113	4	0	0	0	0	24800
MULTI-BIN	105488	22641	2086	108	3	0	0	0	0	24838
MULTINOMIA	107705	18640	3497	426	53	4	0	0	0	22621
MARKOV-BIN	108178	17955	3500	593	88	11	1	0	0	22148
METHERINGH	112758	11478	3804	1454	556	198	61	15	2	17568
COMP-MULTI	113289	8983	6996	232	740	2	78	0	5	17037
MODAL	113777	8957	5716	1082	556	154	69	11	4	16549
CASUALNESS	113674	8967	5863	1170	457	102	44	24	19	16652
METH-C.M.	114332	9338	5044	797	638	94	73	7	4	15994

SCHEDULE 12

	1 ISSUE	2 ISSUES
GLA	4174	6021
HSB	3624	5580
NAT	1831	2793
PSY	1982	2904

NET AND DUPLICATED AUDIENCES

	GLA	HSB	NAT	PSY
GLA	4174	433	38	181
HSB	7365	3624	162	139
NAT	5967	5293	1831	87
PSY	5975	5467	3726	1982

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	114901	9144	5281	523	441	36	1	-0	-0	15425
BINOMIAL	108836	19835	1581	72	2	0	0	0	0	21490
MULTI-BIN	108808	19886	1561	69	2	0	0	0	0	21518
MULTINOMIA	109950	17788	2351	226	14	1	0	0	0	20376
MARKOV-BIN	110656	16588	2667	367	44	5	0	0	0	19670
METHERINGH	114136	11354	3285	1074	345	101	25	5	1	16190
COMP-MULTI	114732	9065	5951	118	431	1	28	0	1	15594
MODAL	114982	9089	5100	772	317	49	15	2	-0	15344
CASUALNESS	114973	9037	5240	745	245	43	19	11	9	15353
METH-C.M.	115508	9581	4262	552	350	47	24	2	1	14818

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 13

	1 ISSUE	2 ISSUES
FOR	2307	3411
AMH	9320	12566
NGE	15842	20841
REA	42616	52989

NET AND DUPLICATED AUDIENCES

	FOR	AMH	NGE	REA
FOR	2307	298	840	1163
AMH	11329	9320	1706	4057
NGE	17309	23456	15842	8596
REA	43760	47879	49872	42616

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	64185	19539	30245	7942	6477	1294	595	31	19	66141
BINOMIAL	41058	51019	27735	8616	1673	208	16	1	0	89268
MULTI-BIN	37891	54509	29300	7524	1023	75	3	0	0	92435
MULTINOMIA	43146	48786	26730	9154	2127	342	37	3	0	87180
MARKOV-BIN	46730	45031	25094	9826	2882	643	107	12	1	83596
METHERINGH	58303	34432	19383	10203	4915	2098	753	205	32	72023
COMP-MULTI	62216	19566	36118	3670	7582	137	975	1	61	68110
MODAL	63835	19800	31412	6653	6578	1130	831	58	29	66491
CASUALNESS	63729	20259	30154	8472	5800	1053	486	160	208	66597
METH-C.M.	61971	25766	27038	6794	6480	1121	990	102	65	68355

SCHEDULE 14

	1 ISSUE	2 ISSUES
HAT	2914	4197
OUT	6372	9626
PLA	16377	22561
TVG	35821	46874

NET AND DUPLICATED AUDIENCES

	HAT	OUT	PLA	TVG
HAT	2914	128	922	914
OUT	9158	6372	1740	2577
PLA	18369	21009	16377	6573
TVG	37821	39616	45625	35821

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	67927	22766	26400	7114	4969	751	368	30	1	62399
BINOMIAL	47754	51082	23906	6393	1069	114	8	0	0	82572
MULTI-BIN	45295	54086	24705	5532	662	43	2	0	0	85031
MULTINOMIA	49558	48939	23323	6883	1398	202	21	1	0	80768
MARKOV-BIN	52714	45257	22277	7667	1968	383	55	5	0	77612
METHERINGH	63066	34699	18029	8689	3798	1456	463	110	15	67260
COMP-MULTI	66058	22749	32037	3224	5524	90	610	0	33	64268
MODAL	67292	23357	27467	6283	4405	955	520	27	9	63034
CASUALNESS	67058	23744	26600	7711	3998	717	274	105	115	63268
METH-C.M.	65478	28162	24707	5891	4685	769	555	55	26	64848

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 15

	1 ISSUE	2 ISSUES
HSB	3624	5580
ESQ	7306	10058
NWK	13690	20029
REA	42616	52989

NET AND DUPLICATED AUDIENCES

	HSB	ESQ	NWK	REA
HSB	3624	327	729	1793
ESQ	10603	7306	2288	3318
NWK	16585	18708	13690	7169
REA	44447	46604	49137	42616

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	64651	22189	27926	8290	5421	1206	582	25	8	65675
BINOMIAL	43179	51149	26509	7851	1453	172	13	1	0	87147
MULTI-BIN	39817	55046	27928	6636	838	59	2	0	0	90509
MULTINOMIA	45317	48903	25430	8395	1929	315	36	3	0	85009
MARKOV-BIN	48652	45156	24201	9106	2565	549	87	10	1	81674
METHERINGH	59880	34502	18947	9716	4549	1883	653	171	26	70446
COMP-MULTI	62119	22884	33841	3976	6508	142	805	1	50	68207
MODAL	65037	20596	30734	6825	4842	1349	818	88	37	65289
CASUALNESS	64434	21579	29267	8695	4684	955	384	150	174	65892
METH-C.M.	62360	27563	26119	6687	5676	1007	784	85	45	67966

SCHEDULE 16

	1 ISSUE	2 ISSUES
ARG	4175	6378
FOR	2307	3411
SIM	10987	16409
WSJ	4732	6485

NET AND DUPLICATED AUDIENCES

	ARG	FOR	SIM	WSJ
ARG	4175	140	782	328
FOR	6342	2307	581	514
SIM	14380	12713	10987	1051
WSJ	8579	6525	14668	4732

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	104220	13600	8775	2628	935	117	43	3	6	26106
BINOMIAL	92008	32741	5097	453	25	1	0	0	0	38318
MULTI-BIN	91752	33188	4964	402	19	1	0	0	0	38574
MULTINOMIA	94915	27822	6350	1091	135	13	1	0	0	35411
MARKOV-BIN	95666	26707	6442	1269	209	29	3	0	0	34660
METHERINGH	102266	17752	6360	2504	967	344	105	25	3	28060
COMP-MULTI	103085	13879	11354	543	1308	7	141	0	9	27241
MODAL	103704	14738	8541	2002	977	251	98	11	3	26622
CASUALNESS	103508	14620	9090	2057	750	168	69	35	25	26818
METH-C.M.	103982	14837	8578	1469	1143	169	129	12	7	26344

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 17

	1 ISSUE	2 ISSUES
GUN	1405	2105
HSB	3624	5580
NAT	1831	2793
SIM	10987	16409

NET AND DUPLICATED AUDIENCES

	GUN	HSB	NAT	SIM
GUN	1405	18	43	370
HSB	5011	3624	162	367
NAT	3193	5293	1831	436
SIM	12022	14244	12382	10987

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	106346	14248	8195	1152	334	45	6	-0	-0	23980
BINOMIAL	98628	27970	3470	246	11	0	0	0	0	31698
MULTI-BIN	98250	28653	3233	184	6	0	0	0	0	32076
MULTINOMIA	99387	26633	3889	390	26	1	0	0	0	30939
MARKOV-BIN	100698	24393	4498	652	77	7	1	0	0	29628
METHERINGH	105020	17907	5202	1588	460	119	25	4	0	25306
COMP-MULTI	106119	14201	9191	206	580	1	28	0	1	24207
MODAL	106234	14581	7963	1074	414	45	13	1	-0	24092
CASUALNESS	106205	14512	8185	1031	306	44	19	11	9	24121
METH-C.M.	106501	15561	6847	855	480	57	24	2	0	23825

SCHEDULE 18

	1 ISSUE	2 ISSUES
BWK	4349	6552
FOR	2307	3411
NAT	1831	2793
WSJ	4372	6485

NET AND DUPLICATED AUDIENCES

	BWK	FOR	NAT	WSJ
BWK	4349	557	625	703
FOR	6099	2307	201	514
NAT	5555	3937	1831	312
WSJ	8018	6165	5891	4372

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	115778	7736	4378	1130	944	186	127	21	29	14548
BINOMIAL	106722	21593	1911	97	3	0	0	0	0	23604
MULTI-BIN	106687	21658	1886	92	3	0	0	0	0	23639
MULTINOMIA	110505	14971	3942	789	109	12	1	0	0	19821
MARKOV-BIN	110192	15653	3576	739	139	23	4	0	0	20134
METHERINGH	115254	9011	3368	1512	699	310	124	40	8	15072
COMP-MULTI	115043	7840	5939	417	900	7	160	0	19	15283
MODAL	115809	7727	4263	1333	736	279	134	32	13	14517
CASUALNESS	115451	7782	4795	1393	581	178	79	38	24	14875
METH-C.M.	116151	7964	4314	882	717	146	122	19	11	14175

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 19

	1 ISSUE	2 ISSUES
COS	4814	6841
HAT	2914	4197
NOB	1787	2601
SAT	2105	3038

NET AND DUPLICATED AUDIENCES

	COS	HAT	NOB	SAT
COS	4814	385	204	207
HAT	7343	2914	218	651
NOB	6397	4483	1787	273
SAT	6712	4368	3619	2105

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	116728	7087	4537	1118	615	187	53	2	-0	13598
BINOMIAL	108820	19847	1584	72	2	0	0	0	0	21506
MULTI-BIN	108783	19917	1556	68	2	0	0	0	0	21543
MULTINOMIA	111385	15191	3256	442	47	4	0	0	0	18941
MARKOV-BIN	111580	15037	3048	555	91	13	2	0	0	18746
METHERINGH	115935	9054	3156	1306	551	220	78	22	4	14391
COMP-MULTI	116299	6989	5953	226	748	2	99	0	8	14027
MODAL	116755	6959	4836	939	544	176	89	20	8	13571
CASUALNESS	116654	6958	4951	1072	467	115	53	28	24	13672
METH-C.M.	117452	7197	4110	723	630	103	94	10	8	12874

SCHEDULE 20

	1 ISSUE	2 ISSUES
LIF	33301	47669
NWK	13690	20029
NEW	3294	4712
SAT	2105	3038

NET AND DUPLICATED AUDIENCES

	LIF	NWK	NEW	SAT
LIF	33301	6565	1538	1130
NWK	40426	13690	1704	852
NEW	35057	15280	3294	569
SAT	34276	14943	4830	2105

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	73795	26091	19700	5900	3370	866	482	95	27	56531
BINOMIAL	55853	49922	19522	4362	609	54	3	0	0	74473
MULTI-BIN	53206	53519	19866	3425	297	13	0	0	0	77120
MULTINOMIA	60615	43463	19064	5768	1210	185	19	1	0	69711
MARKOV-BIN	61645	42472	18499	5915	1465	284	42	4	0	68681
METHERINGH	72746	30490	14811	7128	3230	1320	459	122	19	57580
COMP-MULTI	74420	22195	25514	2848	4543	105	647	1	53	55906
MODAL	74220	24950	21176	5169	3307	715	483	175	130	56106
CASUALNESS	73269	25185	21566	6488	2808	571	235	110	89	57057
METH-C.M.	73936	26476	19898	4950	3751	701	520	61	32	56390

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 21

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	LIF	NOB	DUN	PSY
LIF	33301	47669	33301	827	357	949
NOB	1787	2601	34261	1787	53	149
DUN	680	985	33624	2414	680	46
PSY	1982	2904	34334	3620	2616	1982

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	80231	28542	18794	1856	742	135	26	-0	-0	50095
BINOMIAL	71427	44609	12189	1903	186	12	0	0	0	58899
MULTI-BIN	67436	50956	11277	641	17	0	0	0	0	62890
MULTINOMIA	69056	48415	11570	1202	80	4	0	0	0	61270
MARKOV-BIN	73316	41761	12427	2442	342	35	2	0	0	57010
METHERINGH	77965	35303	12332	3618	894	181	29	3	0	52361
COMP-MULTI	85696	20673	20121	1230	2370	18	209	0	9	44630
MODAL	80531	27226	20486	1232	707	93	47	4	-0	49795
CASUALNESS	80507	27215	20434	1516	529	54	27	20	21	49819
METH-C.M.	82247	28134	15799	2404	1533	98	106	2	4	48079

SCHEDULE 22

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	AMH	BET	ESQ	LAD
AMH	9320	12566	9320	3920	706	3532
BET	22724	30249	28124	22724	1649	6578
ESQ	7306	10058	15920	28381	7306	1050
LAD	16634	22598	22422	32780	22890	16634

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	78900	17434	19721	6290	5289	1338	1155	148	51	51426
BINOMIAL	52519	50550	21286	5122	770	74	4	0	0	77807
MULTI-BIN	51934	51300	21418	4929	683	58	3	0	0	78392
MULTINOMIA	60910	39557	20194	7208	1977	410	66	7	0	69416
MARKOV-BIN	61188	39960	19170	7170	2175	539	107	16	2	69138
METHERINGH	75317	26254	13618	7521	4100	2106	966	359	84	55009
COMP-MULTI	77490	16191	25764	2843	6345	175	1335	4	178	52836
MODAL	78524	18812	18537	6351	5149	1599	1118	151	84	51802
CASUALNESS	78185	18407	19192	7406	4598	1370	669	271	224	52141
METH-C.M.	78521	20159	18217	5221	5313	1160	1349	182	205	51805

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 23

	1 ISSUE	2 ISSUES	NET AND DUPLICATED AUDIENCES MCC	PLA	RED	TIM
MCC	17653	23654	17653	1870	4453	3916
PLA	16377	22561	32160	16377	1519	5465
RED	10473	14412	23673	25331	10473	2385
TIM	20355	27400	34092	31267	28443	20355

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	69784	21084	22751	7845	6091	1626	966	131	48	60542
BINOMIAL	45021	51178	25452	7233	1285	146	10	0	0	85305
MULTI-BIN	44843	51379	25528	7179	1249	138	9	0	0	85483
MULTINOMIA	51638	43774	22454	9425	2468	490	71	0	0	78688
MARKOV-BIN	52698	42600	22545	8901	2757	677	129	18	1	77628
METHERINGH	66443	29927	16342	9037	4801	2353	1009	343	71	63883
COMP-MULTI	69354	18072	30045	3615	7421	223	1427	4	165	60972
MODAL	69590	22081	21993	7753	5794	1799	1063	176	77	60736
CASUALNESS	69583	21440	22280	8926	5351	1584	711	249	198	60743
METH-C.M.	69633	23172	21963	6332	6195	1309	1380	174	168	60693

SCHEDULE 24

	1 ISSUE	2 ISSUES	NET AND DUPLICATED AUDIENCES TIM	FBS	AMH
TIM	20355	27400	20355	594	1909
FBS	1571	2127	21332	1571	104
AMH	9320	12566	27766	10787	9320

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	92560	17667	16868	1884	1309	29	9	131	48	37766
BINOMIAL	79066	41206	8948	1036	68	2	0	0	0	51260
MULTI-BIN	78081	42838	8601	775	31	0	0	0	0	52245
MULTINOMIA	79224	40942	9006	1077	74	3	0	0	0	51102
MARKOV-BIN	83546	34133	10041	2197	363	43	3	18	1	46780
METHERINGH	89524	25995	9766	3569	1145	286	42	343	71	40802
COMP-MULTI	93061	15639	19668	411	1499	1	47	4	165	37265
MODAL	92627	17240	17405	1818	1201	25	10	176	77	37699
CASUALNESS	92747	17183	17221	2026	1005	81	58	249	198	37579
METH-C.M.	92282	20232	14369	1946	1311	140	45	174	168	38044

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 25

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	FWK	PAR	REA	SPO
FWK	16783	20834	16783	3891	6983	1162
PAR	32417	39545	45309	32417	13000	1860
REA	42616	52989	52416	62033	42616	2276
SPO	5916	7541	21537	36473	46256	5916

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	47446	20066	34821	11693	12642	2153	1274	149	82	82880
BINOMIAL	24758	45700	36906	17031	4912	907	105	7	0	105568
MULTI-BIN	23044	46656	38710	16993	4258	614	49	2	0	107282
MULTINOMIA	25373	45778	35902	16727	5234	1131	166	15	1	104953
MARKOV-BIN	29661	42487	32702	16950	6363	1761	352	47	3	100665
METHERINGH	40568	35642	24903	15203	8196	3830	1482	430	72	89758
COMP-MULTI	48043	13820	46225	5226	14273	327	2251	3	157	82283
MODAL	46654	19579	38242	9741	12447	1857	1612	118	77	83672
CASUALNESS	47119	20306	35198	12488	11228	2233	1289	214	246	83207
METH-C.M.	45489	24379	34540	10123	11292	2097	2042	216	148	84837

SCHEDULE 26

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	FWK	PAR	NGE	TVG
FWK	16783	20834	16783	3891	2415	5777
PAR	32417	39545	45309	32417	5366	11253
NGE	15842	20841	30210	42893	15842	4945
TVG	35821	46874	46827	56985	46718	35821

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	45924	20988	34778	10996	12485	3047	1796	258	55	84402
BINOMIAL	23332	44778	37597	18039	5409	1038	125	9	0	106994
MULTI-BIN	22654	45142	38309	18048	5155	914	98	6	0	107672
MULTINOMIA	25126	44491	35220	17431	6156	1577	288	36	2	105200
MARKOV-BIN	28728	41477	32831	17722	7000	2058	441	64	5	101598
METHERINGH	40457	34531	24418	15402	8708	4326	1804	573	107	89869
COMP-MULTI	46830	14923	44168	6182	14709	510	2744	8	252	83496
MODAL	45476	21439	35074	11408	11722	2764	2035	263	145	84850
CASUALNESS	45892	21688	32658	13850	11080	2977	1567	331	279	84434
METH-C.M.	45654	23952	32672	10547	11835	2462	2637	293	274	84672

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

****

SCHEDULE 27

	1 ISSUE	2 ISSUES
REA	42616	52989
PAR	32417	39545
NGE	15842	20841
TVG	35821	46874

NET AND DUPLICATED AUDIENCES

	REA	PAR	NGE	TVG
REA	42616	13000	8586	14093
PAR	62033	32417	5366	11253
NGE	49872	42893	15842	4945
TVG	64344	56985	46718	35821

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	37699	18851	31943	15030	15659	5743	3932	1135	335	92627
BINOMIAL	14048	36084	40549	26038	10450	2684	431	40	2	116278
MULTI-BIN	13518	36077	41267	26355	10246	2473	360	29	1	116808
MULTINOMIA	17575	36426	35893	23544	10242	2860	496	0	0	112751
MARKOV-BIN	19896	34784	34264	23411	11871	4541	1284	249	26	110430
METHERINGH	32649	30550	24436	17831	11930	7198	3765	1564	404	97677
COMP-MULTI	37708	12817	42986	8767	20131	1477	5535	60	846	92618
MODAL	36734	19954	33381	13533	15614	5110	4529	873	598	93592
CASUALNESS	37158	19361	30513	17515	15352	5659	3248	877	639	93168
METH-C.M.	37636	21180	31516	12886	15698	4435	5214	835	926	92690

SCHEDULE 28

	1 ISSUE	2 ISSUES
PAR	32417	39545
FAM	14516	20293
LIF	33301	47669
FOR	2307	3411

NET AND DUPLICATED AUDIENCES

	PAR	FAM	LIF	FOR
PAR	32417	5027	9971	959
FAM	41906	14516	4845	369
LIF	55747	42972	33301	1003
FOR	33765	16454	34605	2307

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	52663	25868	30359	10905	7731	1898	810	71	2	77643
BINOMIAL	32820	49394	32522	12236	2877	433	41	2	0	97506
MULTI-BIN	31062	50798	33974	11905	2324	248	13	0	0	99264
MULTINOMIA	34521	48060	31129	12509	3393	631	78	6	0	95805
MARKOV-BIN	37668	45123	29326	12908	4133	980	168	19	1	92658
METHERINGH	48283	36654	22876	12604	6147	2594	901	233	34	82043
COMP-MULTI	51867	22490	39262	5908	9168	291	1253	3	85	78459
MODAL	52429	26016	31177	9996	7957	1710	959	59	22	77897
CASUALNESS	52522	25960	29834	12397	7139	1579	595	150	148	77804
METH-C.M.	51206	28695	30820	9053	7794	1436	1138	117	67	79120

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 29

	1 ISSUE	2 ISSUES
REA	42616	52989
GOO	15845	22553
TIM	20355	27400
MSB	3624	5580

NET AND DUPLICATED AUDIENCES

	REA	GOO	TIM	MSB
REA	42616	7891	9757	1793
GOO	50570	15845	3336	1793
TIM	53214	32864	20355	1122
MSB	44447	17676	22857	3624

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	57111	22083	28097	11179	8385	1923	1170	140	239	73215
BINOMIAL	32881	49413	32488	12205	2866	431	40	2	0	97445
MULTI-BIN	30653	51373	34151	11664	2230	241	14	0	0	99673
MULTINOMIA	37101	45840	29356	12981	4016	834	136	12	0	93225
MARKOV-BIN	39417	43636	28308	13002	4501	1191	236	33	2	90909
METHERINGH	52436	33530	20712	12096	6542	3180	1322	426	82	77890
COMP-MULTI	54582	21102	36685	5948	9710	413	1709	7	169	75744
MODAL	57325	21267	30210	9317	8253	2118	1374	294	167	73001
CASUALNESS	56841	21677	28902	11871	7625	1992	860	283	270	73485
METH-C.M.	54902	26426	28097	8844	8257	1801	1632	216	151	75424

SCHEDULE 30

	1 ISSUE	2 ISSUES
NGE	15842	20841
RED	10473	14412
NWK	13690	20029
SIM	10987	16409

NET AND DUPLICATED AUDIENCES

	NGE	RED	NWK	SIM
NGE	15842	1685	3417	2917
RED	24630	10473	1625	869
NWK	26115	22538	13690	3227
SIM	23912	20591	21450	10987

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	79726	19533	19286	6056	3584	1614	431	51	45	50600
BINOMIAL	57199	49613	18827	4083	553	48	3	0	0	73127
MULTI-BIN	57122	49718	18836	4057	543	46	2	0	0	73204
MULTINOMIA	63682	40919	18186	5812	1423	262	38	4	0	66644
MARKOV-BIN	64340	40383	17685	5924	1586	342	58	7	1	65986
METHERINGH	76793	27326	13444	6887	3424	1576	634	202	39	53533
COMP-MULTI	78171	19285	24144	2716	4929	127	860	2	93	52155
MODAL	79252	21322	17536	6232	3946	1277	633	95	33	51074
CASUALNESS	79053	20851	18289	6864	3516	1031	427	168	124	51273
METH-C.M.	79560	21669	18107	4699	4371	851	873	101	95	50766

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 31

	1 ISSUE	2 ISSUES
SAT	2105	3038
NEW	3294	4712
HAT	2105	4197

NET AND DUPLICATED AUDIENCES

	SAT	NEW	HAT
SAT	2105	569	651
NEW	4830	3294	548
HAT	3559	4851	2105

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	120977	5037	2607	767	674	208	56	51	45	9349
BINOMIAL	116020	13622	666	17	0	0	0	0	0	14306
MULTI-BIN	116013	13635	661	17	0	0	0	0	0	14313
MULTINOMIA	118764	8601	2506	427	24	3	0	4	0	11562
MARKOV-BIN	118953	8553	2171	513	111	21	4	7	1	11373
METHERINGH	121527	5185	1997	935	442	186	55	202	39	8799
COMP-MULTI	118849	8580	2345	485	56	4	7	2	93	11477
MODAL	121165	4573	2922	842	537	186	101	95	33	9161
CASUALNESS	120943	4615	3221	898	435	141	70	168	124	9383
METH-C.M.	120016	8338	1238	428	199	83	25	101	95	10310

SCHEDULE 32

	1 ISSUE	2 ISSUES
BAR	641	975
FOR	2307	3411
WSJ	4372	6485
BWK	4349	6552

NET AND DUPLICATED AUDIENCES

	BAR	FOR	WSJ	BWK
BAR	641	141	189	165
FOR	2807	2307	514	557
WSJ	4824	6165	4372	703
BWK	4825	6099	8018	4349

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	116595	7564	4171	1046	652	166	84	38	11	13731
BINOMIAL	108737	19918	1596	73	2	0	0	0	0	21589
MULTI-BIN	108672	20038	1549	65	2	0	0	0	0	21654
MULTINOMIA	111744	14558	3374	572	70	7	1	0	0	18582
MARKOV-BIN	111642	14871	3108	586	100	15	2	0	0	18684
METHERINGH	116109	8813	3136	1332	579	239	88	26	5	14217
COMP-MULTI	116050	7557	5534	309	747	4	115	0	11	14276
MODAL	116711	7394	4136	1156	608	194	94	23	10	13615
CASUALNESS	116431	7437	4588	1165	467	128	58	29	19	13895
METH-C.M.	117145	7635	3986	748	594	112	88	12	7	13181

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 33

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	ARG	OUT	SAF	FIE
ARG	4175	6378	4175	823	824	1132
OUT	6372	9626	9724	6372	1443	2176
SAF	5057	6989	8408	9986	5057	1641
FIE	6987	10379	10030	11183	10403	6987

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	108839	8852	6926	2540	1651	954	472	87	6	21487
BINOMIAL	91434	33135	5253	476	27	1	0	0	0	38892
MULTI-BIN	91406	33184	5240	470	26	1	0	0	0	38920
MULTINOMIA	100204	19026	7948	2472	570	90	18	2	0	30122
MARKOV-BIN	98788	21638	7088	2088	556	133	28	5	1	31538
METHERINGH	107760	11559	5127	2754	1553	864	449	199	60	22566
COMP-MULTI	106849	10149	9302	1251	2076	49	538	2	111	23477
MODAL	108248	10412	5815	2741	1644	778	444	163	81	22078
CASUALNESS	107466	10373	7038	2910	1463	577	283	137	73	22860
METH-C.M.	108676	10019	6835	1903	1774	436	495	96	91	21650

SCHEDULE 34

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	BHG	FAM	LHJ
BHG	22724	30249	22724	6124	6578
FAM	14516	20293	31116	14516	4739
LHJ	16634	22598	32780	26411	16634

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	81087	17315	17465	6504	4995	1758	1204	87	6	49239
BINOMIAL	53543	51341	20513	4371	524	33	1	0	0	76783
MULTI-BIN	53387	51555	20530	4317	506	31	1	0	0	76939
MULTINOMIA	64194	36874	19561	7502	1771	378	45	2	0	66132
MARKOV-BIN	67133	34356	17736	7545	2653	746	157	5	1	63193
METHERINGH	79487	22659	12395	7554	4595	2560	1077	199	60	50839
COMP-MULTI	79814	15758	23511	3195	6370	162	1516	2	111	50512
MODAL	81230	18092	16352	6314	5071	1875	1392	163	81	49096
CASUALNESS	80502	17741	7476	4731	1698	951	-0	137	73	49824
METH-C.M.	81703	18064	16655	5470	5517	1367	1549	96	91	48623

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 35

	1 ISSUE	2 ISSUES
BET	22724	30249
FAM	14516	20293
MCC	17653	23654

NET AND DUPLICATED AUDIENCES

	BET	FAM	MCC
BET	22724	6124	6966
FAM	31116	14516	4991
MCC	33411	27178	17653

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	79866	18956	16763	6114	5536	1733	1359	87	6	50460
BINOMIAL	52579	51527	21040	4582	561	37	1	0	0	77747
MULTI-BIN	52435	51722	21060	4531	543	34	1	0	0	77891
MULTINOMIA	63488	36834	19895	7784	1867	408	50	2	0	66838
MARKOV-BIN	66478	34333	17986	7776	2783	798	172	5	1	63848
METHERINGH	79008	22619	12469	7673	4726	2676	1154	199	60	51318
COMP-MULTI	79370	15542	23767	3271	6593	172	1611	2	111	50956
MODAL	80780	17973	16488	6407	5255	1941	1482	163	81	49546
CASUALNESS	80058	17605	17340	7604	4915	1781	1019	137	73	50268
METH-C.M.	81395	17848	16691	5582	5697	1434	1679	96	91	48931

SCHEDULE 36

	1 ISSUE	2 ISSUES
LIF	33301	47669
PLA	16377	22561
NEW	3294	4712

NET AND DUPLICATED AUDIENCES

	LIF	PLA	NEW
LIF	33301	6688	1538
PLA	42990	16377	1193
NEW	35057	18478	3294

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	71507	27248	21045	6404	3394	555	174	87	6	58819
BINOMIAL	54408	51161	20045	4189	492	31	1	0	0	75918
MULTI-BIN	52464	53815	20301	3468	270	8	0	0	0	77862
MULTINOMIA	57000	47464	20023	4995	773	69	3	2	0	73326
MARKOV-BIN	60794	42729	19101	6059	1398	224	21	5	1	69532
METHERINGH	69274	33120	16345	7495	2987	931	174	199	60	61052
COMP-MULTI	71658	23610	28079	2468	4128	38	345	2	111	58668
MODAL	71387	26705	22818	5102	3563	457	293	163	81	58939
CASUALNESS	71130	26572	22860	6282	2813	447	218	137	73	59196
METH-C.M.	70851	28505	21844	4960	3441	479	246	96	91	59475

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 37

	1 ISSUE	2 ISSUES
ESQ	7306	10058
SPO	5916	7541
POP	5565	8098

NET AND DUPLICATED AUDIENCES

	ESQ	SPO	POP
ESQ	7306	782	566
SPO	12440	5916	456
POP	12305	11025	5565

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	107633	11059	9365	1372	824	73	2	87	6	22693
BINOMIAL	96987	29373	3707	249	9	0	0	0	0	33339
MULTI-BIN	96976	29393	3700	248	9	0	0	0	0	33350
MULTINOMIA	98434	26801	4543	509	38	2	0	2	0	31892
MARKOV-BIN	101044	22527	5454	1094	181	24	2	5	1	29282
METHERINGH	105820	15841	5556	2110	746	215	39	199	60	24506
COMP-MULTI	107372	10645	11102	204	952	1	51	2	111	22954
MODAL	107578	10988	9654	1269	740	68	29	163	81	22748
CASUALNESS	107686	10852	9655	1337	634	97	61	137	73	22640
METH-C.M.	108421	11694	7919	1129	988	103	72	96	91	21905

SCHEDULE 38

	1 ISSUE	2 ISSUES
PAR	32417	39545
TVG	35821	46874
REA	42616	52989

NET AND DUPLICATED AUDIENCES

	PAR	TVG	REA
PAR	32417	11253	13000
TVG	56985	35821	14093
REA	62033	64344	42616

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	40701	20814	34543	13536	14655	3893	2186	87	6	89625
BINOMIAL	17629	41857	41411	21850	6485	1027	68	0	0	112697
MULTI-BIN	17519	41905	41545	21854	6434	1005	65	0	0	112807
MULTINOMIA	20885	40915	37176	21307	7944	1855	245	2	0	109441
MARKOV-BIN	25694	37645	33492	20803	9304	2874	514	5	1	104632
METHERINGH	36604	31950	24867	17684	11218	5916	2086	199	60	93722
COMP-MULTI	41282	14582	45056	7883	17559	657	3308	2	111	89044
MODAL	40147	21067	36096	12604	14267	3303	2842	163	81	90179
CASUALNESS	40588	21086	33121	16007	13505	3765	2250	137	73	89738
METH-C.M.	41225	22511	32833	12524	14543	3362	3328	96	91	89101

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 39

	1 ISSUE	2 ISSUES
AMH	9320	12566
FWK	16783	20834
HOT	4038	5826
TST	6448	9263

NET AND DUPLICATED AUDIENCES

	AMH	FWK	HOT	TST
AMH	9320	1425	226	639
FWK	24678	16783	633	734
HOT	13132	20188	4038	245
TST	15129	22497	10241	6448

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	88391	17568	19423	3198	1584	148	16	-0	-0	41935
BINOMIAL	72811	43969	11617	1754	165	10	0	0	0	57515
MULTI-BIN	72349	44691	11523	1620	136	7	0	0	0	57977
MULTINOMIA	72785	44105	11451	1780	190	14	1	0	0	57541
MARKOV-BIN	76494	38449	12063	2760	486	67	7	1	0	53832
METHERINGH	84136	28567	11196	4287	1518	474	122	23	2	46190
COMP-MULTI	87606	17578	22239	710	2082	4	104	0	2	42720
MODAL	88323	17360	20163	2700	1571	149	56	3	1	42003
CASUALNESS	88383	17500	19755	2932	1521	153	53	11	14	41943
METH-C.M.	87286	21676	16745	2415	1840	234	116	11	2	43040

SCHEDULE 40

	1 ISSUE	2 ISSUES
TIM	20355	27400
FAM	14516	20293
LIF	33301	47669

NET AND DUPLICATED AUDIENCES

	TIM	FAM	LIF
TIM	20355	2817	9532
FAM	32054	14516	4845
LIF	44124	42972	33301

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	62779	26191	24133	9126	6502	1075	521	-0	-0	67547
BINOMIAL	41283	52310	27618	7777	1232	104	4	0	0	89043
MULTI-BIN	40613	53093	27891	7536	1106	84	3	0	0	89713
MULTINOMIA	46259	46342	26007	9243	2137	316	22	0	0	84067
MARKOV-BIN	49689	42889	24231	9856	2956	626	79	1	0	80637
METHERINGH	60218	32907	19013	10454	5146	2052	526	23	2	70108
COMP-MULTI	62137	22881	32640	4526	7045	160	938	0	2	68189
MODAL	62369	27059	24015	8715	6355	1187	626	3	1	67957
CASUALNESS	62027	26600	24763	9931	5144	1308	550	11	14	68299
METH-C.M.	61534	27993	25538	7483	5967	1103	708	11	2	68792

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 41

	1 ISSUE	2 ISSUES
WOM	14174	20119
FAM	14516	20293
SIM	10987	16409
NEW	3294	4712

NET AND DUPLICATED AUDIENCES

	WOM	FAM	SIM	NEW
WOM	14174	7093	1051	564
FAM	21597	14516	1136	478
SIM	24110	24367	10987	943
NEW	16904	17332	13338	3294

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1.	2	3	4	5	6	7	8	REACH
OBSERVED	85836	20519	13351	5157	4612	415	361	51	26	44490
BINOMIAL	65486	47063	14798	2659	299	21	1	0	0	64840
MULTI-BIN	65121	47586	14798	2542	262	16	1	0	0	65205
MULTINOMIA	73386	34620	16582	4659	936	146	10	0	0	56940
MARKOV-BIN	72449	37405	14597	4480	1122	230	38	5	0	57877
METHERINGH	84137	24448	11356	5652	2767	1264	508	162	32	46189
COMP-MULTI	85010	18109	20318	2134	3901	89	687	1	77	45316
MODAL	86441	19401	13829	5665	3658	838	415	57	22	43885
CASUALNESS	85681	19333	15626	5629	2712	775	326	139	99	44645
METH-C.M.	85678	20669	15390	3847	3319	668	615	80	60	44648

SCHEDULE 42

	1 ISSUE	2 ISSUES
WOM	14174	20119
FAM	14516	20293
NEW	3294	4712

NET AND DUPLICATED AUDIENCES

	WOM	FAM	NEW
WOM	14174	7093	564
FAM	21597	14516	478
NEW	16904	17332	3294

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	97890	14299	9517	4212	4152	154	105	51	26	32436
BINOMIAL	78098	41748	9299	1105	74	3	0	0	0	52228
MULTI-BIN	77663	42451	9173	983	53	1	0	0	0	52663
MULTINOMIA	86614	26756	14195	2252	483	26	0	0	0	43712
MARKOV-BIN	87605	27567	10538	3424	942	212	37	5	0	42721
METHERINGH	96359	17630	8224	4369	2301	1081	362	162	32	33967
COMP-MULTI	96341	13762	15212	1408	3012	36	555	1	77	33985
MODAL	98387	13734	9323	4660	3705	316	202	57	22	31939
CASUALNESS	97037	14292	11778	4072	2155	652	336	139	99	33289
METH-C.M.	97802	14823	11100	2881	2661	547	512	80	60	32524

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 43

	1 ISSUE	2 ISSUES
FAM	14516	20293
NEW	3294	4712
MCC	17653	23654

NET AND DUPLICATED AUDIENCES

	FAM	NEW	MCC
FAM	14516	478	4991
NEW	17332	3294	581
MCC	27178	20366	17653

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	91154	17597	14713	3898	2697	187	82	51	26	39172
BINOMIAL	73666	44089	10995	1462	109	4	0	0	0	56660
MULTI-BIN	73081	45010	10870	1287	76	2	0	0	0	57245
MULTINOMIA	77818	37064	12805	2318	302	17	0	0	0	52508
MARKOV-BIN	80855	33320	11900	3356	752	129	15	5	0	49471
METHERINGH	88994	23393	10274	4735	2035	727	167	162	32	41332
COMP-MULTI	90561	16382	19241	1069	2784	15	275	1	77	39765
MODAL	91572	16907	14989	3757	2821	189	90	57	22	38754
CASUALNESS	91362	16816	15719	3862	1960	384	219	139	99	38964
METH-C.M.	91226	18940	14130	2896	2502	366	267	80	60	39100

SCHEDULE 44

	1 ISSUE	2 ISSUES
HAT	2914	4197
SAT	2105	3038
LIF	33301	47669

NET AND DUPLICATED AUDIENCES

	HAT	SAT	LIF
HAT	2914	651	1364
SAT	4368	2105	1130
LIF	34851	34276	33301

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	80249	28459	18618	1603	1017	222	161	51	26	50077
BINOMIAL	70184	45758	12430	1801	147	6	0	0	0	60142
MULTI-BIN	66827	51124	11630	726	19	0	0	0	0	63499
MULTINOMIA	69354	47077	12232	1555	104	4	0	0	0	60972
MARKOV-BIN	73680	40446	12925	2804	426	43	2	5	0	56646
METHERINGH	78945	33258	12614	4149	1116	220	24	162	32	51381
COMP-MULTI	83526	22339	21358	1018	1987	5	93	1	77	46800
MODAL	81015	26112	20797	1247	745	247	163	57	22	49311
CASUALNESS	80828	26292	20395	1920	717	104	67	139	99	49498
METH-C.M.	81716	27890	16522	2559	1475	111	54	80	60	48610

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 45

	1 ISSUE	2 ISSUES
HAT	2914	4197
SAT	2105	3038
LIF	33301	47669
TVG	35821	46874

NET AND DUPLICATED AUDIENCES

	HAT	SAT	LIF	TVG
HAT	2914	651	1364	914
SAT	4368	2105	1130	604
LIF	34851	34276	33301	11584
TVG	37821	37322	57538	35821

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	55413	27949	29281	10666	5866	666	437	33	15	74913
BINOMIAL	38197	50665	29402	9750	2021	268	22	1	0	92129
MULTI-BIN	35140	53521	31543	8880	1178	63	2	0	0	95186
MULTINOMIA	38877	49737	29103	10123	2196	285	18	0	0	91449
MARKOV-BIN	42111	46742	27284	10572	2931	592	85	8	0	88215
METHERINGH	51065	38604	22448	11126	4754	1716	497	104	12	79261
COMP-MULTI	55208	24377	38003	4766	7077	142	721	1	32	75118
MODAL	55163	27612	31665	8450	6134	664	411	141	84	75163
CASUALNESS	54936	28176	29868	10912	5298	673	255	99	105	75390
METH-C.M.	53368	31733	29866	7981	5787	925	593	52	21	76958

SCHEDULE 46

	1 ISSUE	2 ISSUES
LIF	15887	22946
REA	22736	28413
SIM	1913	2665
TVG	19216	25165

NET AND DUPLICATED AUDIENCES

	LIF	REA	SIM	TVG
LIF	15887	6763	957	5468
REA	31860	22736	996	7384
SIM	16843	23653	1913	692
TVG	29635	34568	20437	19216

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	19338	13407	16657	8372	6838	2498	1161	60	56	49048
BINOMIAL	9521	21288	20824	11640	4067	909	127	10	0	58865
MULTI-BIN	8772	21408	21797	11927	3751	667	61	2	0	59614
MULTINOMIA	10617	21115	19378	11268	4506	1243	234	26	1	57769
MARKOV-BIN	11908	20330	18366	11067	4793	1520	345	52	4	56478
METHERINGH	17412	17877	13941	9328	5466	2768	1160	366	67	50974
COMP-MULTI	19580	9585	23154	5224	8450	564	1649	12	168	48806
MODAL	19589	12648	18014	7441	6969	2075	1391	160	99	48797
METH-C.M.	18763	13694	18338	7257	6925	1673	1424	189	124	49623

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 47

	1 ISSUE	2 ISSUES	NET AND DUPLICATED AUDIENCES FAM	MCC	WOM	LAD
FAM	12357	16771	12357	4552	6457	4369
MCC	14594	19064	22399	14594	4466	6247
WOM	12590	17830	18490	22718	12590	4492
LAD	14182	19181	22170	22529	22280	14182

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	29553	11404	9932	5677	5644	2592	2043	927	603	38833
BINOMIAL	11893	23253	19890	9722	2970	581	71	5	0	56493
MULTI-BIN	11878	23261	19906	9723	2965	578	70	5	0	56508
MULTINOMIA	19984	17577	13758	9623	4769	1844	651	162	19	48402
MARKOV-BIN	18749	18919	14593	8926	4492	1873	637	167	30	49637
METHERINGH	29063	13207	8642	6113	4393	3101	2077	1240	550	39323
COMP-MULTI	28978	8176	15616	4076	7154	831	2663	82	812	39408
MODAL	29926	11004	10564	5526	4659	2714	2168	1074	750	38460
METH-C.M.	30451	10084	11240	4895	5692	1963	2525	673	863	37935

SCHEDULE 48

	1 ISSUE	2 ISSUES	NET AND DUPLICATED AUDIENCES TIM	USN	NWK	SAT
TIM	8627	11181	8527	1060	1653	455
USN	3076	4633	10643	3076	833	187
NWK	4743	6670	11717	6986	4743	286
SAT	861	1258	9033	3750	5318	861

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	51084	6553	6941	2179	964	324	242	78	21	17302
BINOMIAL	40540	21906	5178	700	59	3	0	0	0	27846
MULTI-BIN	40219	22421	5092	611	42	2	0	0	0	28167
MULTINOMIA	44128	16342	5931	1599	326	52	7	1	0	24258
MARKOV-BIN	44241	16393	5643	1617	394	82	14	2	0	24145
METHERINGH	49848	9950	4403	2203	1114	537	232	82	18	18538
COMP-MULTI	49943	7544	8272	724	1553	26	288	0	35	18443
MODAL	51156	6881	6562	1762	1260	417	264	52	31	17230
METH-C.M.	51032	7881	5944	1445	1406	278	316	40	42	17354

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 49

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	LIF	REA	SIM	TVG
LIF	8748	12577	8748	3689	708	2925
REA	11683	14468	16742	11683	771	4100
SIM	1392	1926	9432	12304	1392	476
TVG	9090	11976	14913	16673	10006	9090

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	6723	5974	7735	4325	3859	1395	718	37	54	24097
BINOMIAL	3061	8195	9599	6425	2688	720	120	12	0	27759
MULTI-BIN	2762	8134	10024	6675	2579	574	68	4	0	28058
MULTINOMIA	3304	8386	9174	6078	2778	878	192	26	2	27516
MARKOV-BIN	3913	8104	8602	5960	2915	1027	255	41	3	26907
METHERINGH	6014	7611	6727	4893	3036	1597	685	218	40	24806
COMP-MULTI	6999	3922	10755	2979	4668	413	971	10	102	23821
MODAL	6831	5597	8472	3883	3803	1221	836	113	65	23989
METH-C.M.	6468	5865	8683	4014	3792	1014	803	114	67	24352

SCHEDULE 50

NET AND DUPLICATED AUDIENCES

	1 ISSUE	2 ISSUES	FAM	MCC	WOM	LAD
FAM	6919	9215	6919	2671	3879	2781
MCC	7851	10041	12099	7851	2713	3709
WOM	7185	10239	10225	12323	7185	2764
LAD	8071	10803	12209	12213	12492	8071

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	10582	4948	5120	3170	3321	1504	1100	644	431	20238
BINOMIAL	3304	8510	9591	6176	2486	640	103	9	0	27516
MULTI-BIN	3299	8511	9598	6179	2484	638	102	9	0	27521
MULTINOMIA	6351	7400	6660	5272	3068	1363	534	152	21	24469
MARKOV-BIN	5899	7732	7139	5085	2920	1365	511	145	27	24921
METHERINGH	10319	6054	4374	3300	2493	1837	1280	795	369	20501
COMP-MULTI	10339	3487	7651	2346	4123	630	1648	74	522	20481
MODAL	10688	5050	5303	2998	2672	1606	1335	664	504	20132
METH-C.M.	10738	4646	5750	2748	3254	1231	1507	441	506	20082

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 51

	1 ISSUE	2 ISSUES
TIM	5527	7060
USN	1950	2896
NWK	3011	4250
SAT	668	962

NET AND DUPLICATED AUDIENCES

	TIM	USN	NWK	SAT
TIM	5527	689	1100	354
USN	6788	1950	560	175
NWK	7438	4401	3011	203
SAT	5841	2443	3476	668

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	20073	3733	4336	1547	707	206	127	78	13	10747
BINOMIAL	14430	11486	4000	796	99	8	0	0	0	16390
MULTI-BIN	14193	11822	4008	718	74	4	0	0	0	16627
MULTINOMIA	16118	9157	3923	1257	301	56	8	1	0	14702
MARKOV-BIN	16279	9025	3809	1268	345	77	14	2	0	14541
METHERINGH	19297	5844	2836	1481	764	371	160	55	12	11523
COMP-MULTI	19485	4153	5271	558	1099	26	204	0	24	11335
MODAL	20071	3967	4232	1176	851	278	176	43	25	10749
METH-C.M.	19970	4557	3829	1001	984	199	222	28	29	10850

SCHEDULE 52

	1 ISSUE	2 ISSUES
LIF	4292	5982
REA	5996	7046
SIM	678	1060
TVG	3510	4877

NET AND DUPLICATED AUDIENCES

	LIF	REA	SIM	TVG
LIF	4292	2047	345	1350
REA	8241	5996	403	1874
SIM	4625	6271	678	263
TVG	6452	7632	3925	3510

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	2858	2464	3381	1973	1871	751	372	31	15	10858
BINOMIAL	1183	3392	4255	3050	1367	392	70	7	0	12533
MULTI-BIN	1026	3341	4484	3204	1313	307	38	2	0	12690
MULTINOMIA	1301	3498	4056	2858	1397	475	112	17	1	12415
MARKOV-BIN	1559	3396	3805	2792	1450	543	144	25	2	12157
METHERINGH	2496	3248	2968	2242	1451	801	363	123	24	11220
COMP-MULTI	2779	1774	4687	1521	2179	244	472	7	53	10937
MODAL	2937	2228	3786	1787	1780	643	458	68	28	10779
METH-C.M.	2684	2479	3832	1855	1823	518	421	65	39	11032

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 53

	1 ISSUE	2 ISSUES
FAM	3247	4470
MCC	4130	5270
WOM	3157	4615
LAD	4230	5448

NET AND DUPLICATED AUDIENCES

	FAM	MCC	WOM	LAD
FAM	3247	1418	1710	1542
MCC	5959	4130	1371	2211
WOM	4694	5916	3157	1475
LAD	5935	6149	5912	4230

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	4197	2291	2189	1308	1766	725	720	247	276	9519
BINOMIAL	1117	3290	4240	3122	1437	423	78	8	0	12599
MULTI-BIN	1106	3288	4255	3131	1433	418	76	8	0	12610
MULTINOMIA	2403	3049	2940	2502	1601	769	328	105	17	11313
MARKOV-BIN	2198	3164	3178	2455	1526	771	312	95	19	11518
METHERINGH	4122	2614	1972	1543	1207	924	672	441	221	9594
COMP-MULTI	4059	1517	3375	1188	2010	377	849	53	289	9657
MODAL	4332	2126	2357	1436	1294	818	656	393	303	9384
METH-C.M.	4076	2091	2675	1388	1599	654	744	246	244	9640

SCHEDULE 54

	1 ISSUE	2 ISSUES
TIM	3613	4897
USN	1519	2284
NWK	2065	2862
SAT	721	1023

NET AND DUPLICATED AUDIENCES

	TIM	USN	NWK	SAT
TIM	3613	617	796	397
USN	4515	1519	471	181
NWK	4882	3113	2065	258
SAT	3937	2059	2528	721

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	6474	2537	2553	1133	601	241	80	78	21	7242
BINOMIAL	3942	5319	3140	1059	223	30	3	0	0	9774
MULTI-BIN	3811	5449	3218	1023	191	22	1	0	0	9905
MULTINOMIA	4688	4587	2730	1192	395	98	19	3	0	9028
MARKOV-BIN	4757	4519	2708	1191	406	109	22	3	0	8959
METHERINGH	6255	3279	1923	1119	621	317	141	50	11	7461
COMP-MULTI	6389	2247	3332	588	899	48	186	1	25	7327
MODAL	6543	2545	2539	937	665	261	160	43	22	7173
METH-C.M.	6353	2777	2603	855	745	181	159	25	17	7363

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 55

	1 ISSUE	2 ISSUES
LIF	1339	1871
REA	1683	1894
SIM	844	1385
TVG	1063	1150

NET AND DUPLICATED AUDIENCES

	LIF	REA	SIM	TVG
LIF	1339	691	427	467
REA	2331	1683	407	606
SIM	1756	2120	844	263
TVG	1935	2140	1644	1063

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	566	420	943	656	521	221	153	121	79	3114
BINOMIAL	141	568	1001	1007	634	255	64	9	1	3539
MULTI-BIN	132	558	1012	1027	637	247	59	8	0	3548
MULTINOMIA	184	649	959	869	588	294	105	28	4	3496
MARKOV-BIN	235	630	900	864	600	307	113	28	4	3445
METHERINGH	485	682	708	629	496	346	207	98	29	3195
COMP-MULTI	457	396	1054	568	741	190	224	15	37	3223
MODAL	572	503	845	554	538	317	220	97	34	3108
METH-C.M.	397	567	889	687	609	277	178	54	22	3283

SCHEDULE 56

	1 ISSUE	2 ISSUES
FAM	143	249
MCC	144	225
WOM	149	197
LAD	162	215

NET AND DUPLICATED AUDIENCES

	FAM	MCC	WOM	LAD
FAM	143	15	59	30
MCC	272	144	19	39
WOM	233	274	149	29
LAD	275	267	282	162

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	3064	301	173	82	23	12	25	-0	-0	616
BINOMIAL	2641	895	133	11	1	0	0	0	0	1039
MULTI-BIN	2641	895	133	11	1	0	0	0	0	1039
MULTINOMIA	2858	541	202	70	8	2	0	0	0	822
MARKOV-BIN	2822	606	185	51	12	3	1	0	0	858
METHERINGH	3057	332	142	73	39	21	10	4	1	623
COMP-MULTI	3003	339	244	35	45	1	10	0	2	677
MODAL	3076	291	161	75	43	21	10	3	1	604
METH-C.M.	2993	405	186	47	31	10	6	2	1	687

DATA BASE AND ESTIMATED FREQUENCY DISTRIBUTIONS

SCHEDULE 57

	1 ISSUE	2 ISSUES
TIM	1537	2094
USN	629	782
NWK	978	1404
SAT	100	130

NET AND DUPLICATED AUDIENCES

	TIM	USN	NWK	SAT
TIM	1537	345	507	63
USN	1821	629	250	32
NWK	2008	1357	978	46
SAT	1574	697	1032	100

OBSERVED VERSUS ESTIMATED FREQUENCY DISTRIBUTIONS

FREQUENCY	0	1	2	3	4	5	6	7	8	REACH
OBSERVED	1042	649	1062	277	472	90	75	14	-0	2638
BINOMIAL	502	1136	1124	635	224	51	7	1	0	3178
MULTI-BIN	438	1150	1208	654	196	32	3	0	0	3242
MULTINOMIA	576	1116	1033	616	252	72	14	2	0	3104
MARKOV-BIN	629	1087	992	603	263	84	19	3	0	3051
METHERINGH	922	960	754	507	298	151	63	20	4	2758
COMP-MULTI	1021	552	1219	307	447	35	89	1	10	2659
MODAL	989	769	958	394	355	111	89	8	7	2691
METH-C.M.	982	753	983	405	371	93	76	10	7	2698

BIBLIOGRAPHY

Aaker, D.A. "A Probabilistic Approach to Industrial Media Selection." Journal of Advertising Research, Vol. 8, September 1968, pp. 46-54.

_____. "ADMOD: An Advertising Decision Model." Journal of Marketing Research, Vol. 12, February 1975, pp. 37-45.

Adam, Daniel et al. "Problems Posès par la Cumulation d'Audience dans le Temps." Etude I.R.E.P. N.7, Paris 1962.

Advertising Research Foundation. A.R.F.'s Total Audit of Simmons, ARF, Vol. I, II, III, IV, New York, August 1975.

Advertising Research Foundation. "Towards Better Media Comparisons." New York: Advertising Research Foundation Audience Concept Committee, 1961.

"Agency Researchers Stumped by Big Swing in New Simmons Numbers." Media Decisions, December 1974, pp. 28-29.

Agostini, J.M. "How to Estimate Unduplicate Audiences." Journal of Advertising Research, Vol. 1, March 1961, pp. 11-14.

_____. "Analysis of Magazine Accumulative Audience." Journal of Advertising Research, Vol. 2, October 1962, pp. 24-27.

_____. "Les Questions Directes sur les Habitudes de Lecture sont-elles Vraiment si Mauvaises?" XV Esomar Conference 1962.

_____. "The Case for Direct Questions on Reading Habits." Journal of Advertising Research, June 1964.

Alpert, L. "Measuring Effects Through Simulation." Proceedings of the 18th Annual Conference of the A.R.F., November 1972, pp. 48-50.

American Market Research Bureau. "Questions about T.G.I." New York: American Market Research Bureau, April 1972.

Anscombe, F.J. "Sampling Theory of the Negative Binomial and Logarithmic Distributions." Biometrica, Vol. 37, 1950.

Axelrod, J.N. Advertising Research Foundation Guide to Copy Research. New York: Advertising Research Foundation, September 1971.

Banks, S., Mandasky, A. "Estimation of Multimagazine Readership." Journal of Business, Vol. 31, July 1958, pp. 235-242.

Banks, S. "Patterns of Daytime Viewing Behavior." Proceedings of the American Marketing Association, June 1967, pp. 139-142.

Barlett, M.S. "Contingency Table Interactions." Journal of the Royal Statistical Society, Supplement 2, 1935, pp. 248-252.

Bass, F.M., Lansdale, R.T. "An Exploration of Linear Programming in Media Selection." Journal of Marketing Research, May 1966, pp. 279-188.

Bass, F.M., Pessemier, E.A., Tigert, D.J. "A Taxonomy of Magazine Readership Applied to Problems in Marketing Strategy and Media Selection." Journal of Business, Vol. 42, July, 1969, pp. 337-363.

Beale, E.M.L., Hughes, P.A.B., Broadbent, S.R. "A Computer Assessment of Media Schedules." Operational Research Quarterly, Vol. 17, December 1966, pp. 381-411.

Benson, P.H. "Bivariate Normal Distribution to Calculate Media Exposure." Journal of Advertising Research, Vol. 9, 1969, pp. 41-47.

_____. "Segmentation Analysis of Media Exposure: Toward a Theory of Learning Opportunity." Convention Proceedings, American Psychological Association, 1971.

Bernay, E.K. "Life Style Analysis as a Basis for Media Selection." In C.W.King and D.J. Tigert, editors, Attitude Research Reaches New Heights, 1971, pp. 189-195.

Bernstein, A. "Computer Simulation of Media Exposure." A Report of the Sixth Meeting of the A.R.F. Operation Research Discussion Group, New York, 1961.

Bertier, P., Du Jeu, P. "Simulation des Comportements Aleatoires de Lecture: Le Modele SCAL." Metra, Vol. 6, No. 4, pp. 647-659.

Biecheler, A. "Essai de Determination du Profil des Lecturs par des Methodes d'analyse Multidimensionnelle Appliquee a l'Etude du CESP sur les Domaines l'Interet." Bulletin de l'IREP, No. 23, September 1972, pp. 5-35.

Bogart, Leo, ed. Psychology in Media Strategy. American Marketing Association, Chicago 1966.

Bower, J. "New Audiences of U.S. and Canadian Magazines: Seven Tests of Agostini's Formula." Journal of Marketing Research, Vol. 3 March 1963, pp. 13-20.

Boyd, H.W., Claycamp, H.J., McClelland, C.N. "Media Models for the Industrial Goods Advertiser: A Do it Yourself Opportunity." Journal of Marketing, April, 1970, pp. 23-27.

Broadbent, S.R. "Regularity of Reading." Journal of Marketing Research, August 1964, pp. 50-58.

_____. "Media Planning and Computers by 1970." Silver Medal Winning Paper, The Thomson Medals and Awards for Media Research, 1965, London, pp. 61-86.

Broadbent, S.R., Segnit, S. "Response Functions in Media Planning." The Thomson Medal and Awards for Media Research, London 1967, pp. 35-86.

_____. "Beyond Cost per Thousand- An Examination of Media Weights." The Thomson Medals and Awards for Advertising Research, 1968, pp. 105-140.

Brown, I.O., Lessler, R.S., Weilbacher, W.M. "Advertising Media: Creative Planning in Media Selection." New York: Ronald Press, 1967.

Brown, D.B., Warshaw, M.R. "Media Selection by Linear Programming." Journal of Marketing Research, February 1965, pp. 83-88.

Caffyn, J.M., Sagovsky, M. "Net Audiences of British Newspapers: A Comparison of the Agostini and Sainsbury Methods." Journal of Advertising Research, Vol. 3, March 1963, pp. 21-24.

Catry, B., Chevalier, M. "The Evolution of French Media Models." Journal of Advertising Research, Vol. 13, June 1973, pp. 19-26.

Chandon, J.L. "M.T.M.: A Media Timing Model." C.E.S.A.M. Working Paper, Louvain, 1973.

_____. "Le Point sur les Modèles d'Accumulation d'Audience." La Revue Francaise du Marketing, Cahier 53, 4° Timestre 1974, pp. 51-76.

_____. Report on the performance of SUPERCUME, a network TV reach and frequency model , Interactive Market Systems Inc., New York 1974.

Charnes, A., Cooper, W.W., Learner, D.B., Snow, E. "Media Mixing for Distribution of Frequencies with Beta like and Lognormal Distributions of Frequencies." Management Science Report, No. 189, Carnegie Millon University, Pittsburgh, January 1968.

Chatfield, C. and G. Goodhart. "The Beta Binomial Model for Consumer Purchasing Behavior." Applied Statistics, Vol. 19, Part 3, pp. 240-250, 1970.

Chinsky, J.M., Wells, W.D. "Effects of Competing Messages: A Laboratory Simulation." Journal of Marketing Research, May 1965, pp.141-148.

Claycamp, H.J. and McClelland, C.W. "Estimating Reach and the Magic of K." Journal of Advertising Research, Vol. 8, No. 2, 1968, pp. 44-51.

Coen, Fobert G. "1975 Ad Volume may top $28 billion following 5.9% increase for 1974." Advertising Age, December 16, 1974, p.1.

The Concept of Effective Reach. Unauthored internal memo, J.W. Thomson Agency, November, 1973.

Corlett, T., Pretty, B.J., Rotham, L.J. Thomson Award for Media Research, London 1962.

Corlett, T., Osborne, D.W. The Development of Reaching Frequency Scales. I.P.A., London 1966.

COSMIC- A Comprehensive Press Planning System. Foote, Cone and Belding, London, September 1970.

Day, R.L. "Linear Programming in Media Selection." Journal of Advertising Research, Vol. 2, June 1962, pp.40-44.

Deming, M.E. and Stephan, F.F. "On a least square adjustment of a sampled frequency table when the expacted marginals are known." Annals of Mathematical Statistics, Vol. 11, pp. 427-444, 1970.

Diamond, D.S. "A Quantitative Approach to Magazine Advertisement Format Selection." Journal of Marketing Research, Vol. 5, November 1968.

Damaracki, J. BLAZT, A USER'S GUIDE TO THE LEO BURNETT COMPANY SCHEDULING MODEL. Internal Memo, Leo Burnett Co., Chicago 1970.

Durand, J. "Recherche des Données Necessaires au Fonctionnement des Modèles.", Seminaire I.R.E.P.- La Recherche Operationnelle Appliquèe à l'Elaboration des Plans Media, 1966.

_____. "Advertising and Time: Growth and Decay." The Marcel Dassault Awards, Paris 1969, pp. 47-93.

Dynamo-Media Scheduling Suit of Programs. University of London, Atlas Computing Service.

Ebbinghaus, H. Uber das Gedachtnis. Dunker and Humbolt, Leipzig 1885.

Ehrenberg, A.S.C., Twynam, W.A. "On Measuring Television Audiences." Journal of the Royal Statistical Society, Serie A-Vol. 130, 1967, pp. 1-59.

_____. Repeat Buying Theory and Applications. North Holland Publishing Co., Amsterdam, 1972.

Ellis, D.M. "Building up a Sequence of Optimum Mdeia Schedules." Operational Research Quarterly, Vol. 17, December 1966, New York, pp. 413-423.

Engel, J., Warshaw, M.R. "Allocating Advertising Dollar by Linear Programming." Journal of Advertising Research, Vol. 4, September 1966, pp. 42-48.

Engelman, F.L. "An Empirical Formula for Audience Accumulation." Journal of Advertising Research, Vol. 5, June 1965, pp. 21-28.

Federman, S.P. "Filling the Syndicated Data Gap." Proceedings of the 19th Annual Conference of the A.R.F., New York, 1974.

Feller, W. An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed., Wiley, New York, 1969.

Fienberg, S.E. The Analysis of Cross Classified Data. Working Paper, Department of Statistics, University of Chicago, 1973.

Frank, R.E. "Television Program Audience Similarities: A Taxonomic Analysis." Working Paper, University of Pennsylvania, December, 1967.

Frank, R.E., Becknell, J.C., Cleckey, J.D. "Television Program Types." Journal of Marketing Research, Vol. 8, May 1971, pp. 204-211.

Frankel, L.R. and Solomon, D. Look Audience- 1962, A National Study, Audits and Surveys Co. Inc. New York 1962.

Frankel, L.R. "Some Recent Developments in Market Research Methods." XV Esomar Conference, 1962.

Frankel, L.B. "Mass Media: The Process of Communication." WBM/IFM Congress on Communication Research, Hamburg 1963.

Friedman, L. "Calculating TV Reach and Frequency." Journal of Advertising Research, Vol. II, August 1971.

Fry, C. "The Concept of Casualness." Feedback, No. 2, April 1973.

Gallup and Robinson, Inc. An Analysis of Television Commercial Build-Up/Wear-Out. Princeton, March 1967.

Gensch, D. Advertising Planning: Mathematical Models in Advertising Media Planning, Elsevier Scientific Publishing Co., Amsterdam 1973.

_____. "Computer Models in Advertising Media Selection." Journal of Marketing Research, Vol. 5, November 1968, pp. 414-424.

_____. "A Computer Simulation Model for Selecting Advertising Schedules." Journal of Marketing Research, Vol. 7, May 1969, pp. 203-214.

_____. "Media Factors, A Review Article." Journal of Marketing Research, Vol. 7, May 1970, pp. 216-225.

Gerhold, P.E. "How to Measure Magazine Audiences." 19th Annual Conference, 1973 Proceedings. Advertising Research Foundation, pp. 47-49.

Goodhart, G.J., Ehrenberg, A.S.C. "Duplication of Television Viewing Between and Within Channels." Journal of Marketing Research, Vol. 6, May 1969, pp. 169-178.

Graham, I. Encyclopedia of Advertising. Fairchild, 2d edition, New York 1969.

Grass, R.G. "Saturation Effects of Advertising." Proceedings of the 14th Annual Conference of the A.R.F., New York, October 1968, pp. 20-28.

Grass, R.C., Wallace, W.H. "Saturation Effects of TV Commercials." Journal of Advertising Research, Vol. 9, June 1969, pp. 3-8.

Greene, J., Stock, J.S. "Advertising Reach and Frequency in Magazines: New Information for Evaluating Magazine Advertising." Readers Digest Association, Inc., New York 1967.

Greene, J.D. "Personal Media Probalilities." Journal of Advertising Research, Vol. 10, October 1970, pp. 12-18.

Greene, J., Stock, J.S. "A Rate Frequency Model of Behavior." Journal of Advertising Research, Vol. 11, August 1971, pp. 9-19.

Greenberg, A., Suttoni, G. "Television Commercial Wearout." Journal of Advertising Research, Vol.13, October 1973, pp. 47-54.

Greenwood, M. and Yule, G.U. "An Inquiry into the Nature of Frequency Distribution Representative of Multiple Happenings." Journal of the Royal Statistical Society, Vol. 83, 1920.

Gross, I. "The Creative Aspects of Advertising." Sloan Management Review, Vol. 14, Fall 1972, pp. 83-110.

Grueber, A. "Top of Mind Awareness and Share of Families: An Observation." Journal of Marketing Research, Vol. 7, May 1969, pp.227-231.

Haley, R.I. "Do We Really Know What We are Doing?" Proceedings of the American Marketing Association Conference, 1971, pp. 217-219.

Header, R.S. "Distribution Theory of Net Cumulative Audience Coverage and Exposure Frequency in Media Planning." Proceedings of the 1971 Conference, American Marketing Association, pp. 230-234.

Hensel, J., Steventhal, B. Physiological Measures of Information Receptivity: Galvanic Skin Response and Pupil Dilation. Working Paper, Northwestern University 1974.

Hofmans, P. "Measuring the Cumulative Net Coverage of any Combination of Media." Journal of Marketing Research, Vol. 3, August 1966, pp. 269-278.

Hofmans, P. "Measuring the Cumulative Net Coverage of any Combination of Media." Journal of Marketing Research, Vol. 3, August 1966, pp. 269-278.

Hughes, M. "A Model of the Evolution of Advertising Effectiveness." The Marcel Dassault Awards, Paris 1969, pp. 9-42.

Hyett, G.P. "The Measurement of Readership." London School of Economics Paper, 1958, London.

Hulks, B. and Thomas, S.G. "PREFACE, a simple model for the prediction of television coverage and frequency distribution." XXVI ESOMAR/WAPOR Congress, Budapest 1972.

INTERMARCO Limited. Advertising Planning by Media Planex, 1971.

Institute of Practitioner in Advertising. Survey of Press Readership, London 1939.

Ishii, G. and Hayakawa, R. "On the compound Binomial Distribution." Annals of the Institute of Statistics Mathematic, Vol. 12,1960 pp. 69-80, Tokyo.

Johnson, N.L., Kotz, S. Continuous Multivariate Distributions. Wiley and Sons 1972.

Johnson, N.L., Kotz, S. Discrete Distributions. Houghton Mifflin, Boston 1969.

Joyce, T. "Measuring the Audience of Magazines." American Market Research Bureau, April 1972, New York.

Joyce, T., Bird, M. "The use of panels in readership survey." The Thomson Awards for Media Research, 1966, pp. 49-70.

Kaatz, R.B. "Improving Agostini Formula for New Audience." Letter to the Editor, Journal of Advertising Research, Vol. 3, September 1963, pp. 43-44.

Kendall and Stuart. Advanced Theory of Statistics, Vol. 1, Griffin, London 1958.

King. C.W., and Summers, J.O. "Attitudes and Media Exposures." Journal of Advertising Research, Vol. II, February 1971, pp. 26-32.

Klotz, J. "Statistical Inference in Bernouilli Trials with Dependence." The Annals of Statistics, 1973, Vol. 1, No. 2, pp. 373-379.

Kotler, Philip. "Toward an Explicit Model for Media Selection." Journal of Advertising Research, Vol. 4, March 1964, pp. 34-41.

Kuhn, W. "Net Audiences of German Magazines: Seven Tests of Agostini's Formula." Journal of Advertising Research, Vol. 3, March 1963, pp. 13-20.

Kwerel, S.M. Estimating the Unduplicated Audience of a Combination of Media Vehicles- Integrated Theory and Estimation Method, Ph.D Dissertation, Columbia University, 1964.

_____. "Estimating Unduplicated Audience and Exposure Distribution." Journal of Advertising Research, Vol. 9, June 1969, pp. 46-53.

Lee, A.M., Burkart, A.J. "Some Optimization Problems in Advertising Media Planning." Operational Research Quarterly, Vol. 11, No. 3, September 1960, pp. 113-122.

Lee, A.M. "Decision Rules for Media Scheduling: Static Campaigns." Operational Research Quarterly, Vol. 13, No. 3, September 1962, pp. 229-242.

_____. "Decision Rules for Media Scheduling: Dynamic Campaigns." Operational Research Quarterly, Vol. 14, December 1963, pp. 365-372.

Leheman, D.R. "Television Show Preference: Application of a Choice Model." Journal of Marketing Research, February 1971, pp. 47-58.

Liebman, L. "Modal Reach and Frequency: An Improved Estimation Method." 17th Annual Conference of the A.R.F., November 1971.

_____. "Qualitative Data in Media Planning: New Developments at I.M.S." Proceedings of the 19th Annual Conference of the A.R.F., New York 1974.

"Continuing Study of Magazine Audiences." Report No. 1, Life, New York, December 1938.

Little, J.D.C., and Lodish, L.M. "A Media Planning Calculus." Operation Research, January 1969, pp. 3-35.

Little, J.D.C. "Models and Managers: The Concept of a Decision Calculus." Management Science, Vol. 16, April 1970, B466-485.

Lodish, L.M. "Considering Completion in Media Planning." Management Science, Vol. 17, February 1971, B293-306.

_____. "Exposure Interaction Among Media Schedules." Journal of Advertising Research, Vol. 13, April 1973, pp. 31-34.

Lucas, D.B. "Mathematical Programming for Better Media Selection." Address to the American Association of Advertising Agencies, November 1961, New York.

Lucas, D.B., Britt, S.H. Measuring Advertising Effectiveness, McGraw-Hill, New York, 1963.

Maloney, J.C. "Copy Testing: What Course is it Taking." Proceedings of the 9th Annual Conference of the A.R.F., New York, 1964.

_____. "Is Advertising Believability Really Important?" Journal of Marketing, Vol. 27, October 1963, pp. 1-8.

Marc, M. "Etude de la Couverture Realisee par Plusieurs Revues Techniopus sur une Clientele Speciale." I.R.E.P., Etude No. 6 Bis, Paris, Novembre 1962.

_____. "Net Audiences of French Business Papers: Agostini's Formula Applied to Special Markets." Journal of Advertising Research, Vol. 3, March 1963, pp. 26-29.

_____. "Combining Simulation and Panel Data to Obtain Reach and Frequency." Journal of Advertising Research, Vol. 8, No. 2 June 1968, pp. 11-16.

Marder, E. "How good is the Editorial Interest Method of Measuring Magazine Audiences?" Journal of Advertising Research, Vol. 7, March 1967.

Mauldon, J.G. "A Generalization of the Beta Distribution." Annals of Mathematical Statistics, Vol. 30, pp. 509-520, 1959.

"Magazine Research: Who is Suing Who?" Media Decisions, February 1975.

Menke, W.W. "Linear Programming Applied to Marketing Decision Making." Proceedings of the American Marketing Association, June 1968, pp. 245-252.

Metheringham, R.A. "Measuring the Net Cumulative Coverage of a Print Campaign." Journal of Advertising Research, Vol. 4, December 1964, pp. 23-28.

_____. "Models, Media Measurement, and the Metheringham Formula." Proceedings of the 17th Annual A.R.F. Conference, November 1971.

Moran, W.T. "Practical Media Decision and the Computer." Journal of Marketing, Vol. 27, July 1963.

Morgensztern, A. "Frequentation de la Presse, Generalisation, Effet Publicitaire." Etude No. 19, I.R.E.P., Avril 1970, Paris.

Mosimann, J.E. "On the Compound Multinomial Distribution, the Multivariate β Distribution, and Correlations Among Proportions." Biometrica, Vol. 49, 1962, pp. 65-82.

Papazian, E. "Media Research Jumps from the Frying Pan into the Fire." Media Decisions, December 1974, pp. 12-14.

Politz (Alfred) Research, Inc. A Study of the Household Accumulative Audience of Life, New York, 1952.

_____. A Study of Four Media, Their Accumulative and Repeat Audiences, A Politz Research, New York, Time Inc., 1953.

_____. A 12 Month Study of Better Homes and Gardens Readers, Meredith Publishing Co., Des Moines.1956.

_____. Life Study of Consumer Expenditures, 7 Volumes, Time Inc., New York 1958.

Politz, Alfred Media Studies. Ad Page Exposure in 4 Magazines, New York 1960.

_____. Main Report: 1964 Politz Magazine Study- Audiences of 11 Magazines, New York, February 1965.

_____. An Experimental Study Comparing Magazine Audiences as Determined by Two Questioning Procedures, New York 1967.

Ramond, C. "Toward Valid Intermedia Comparisons of Response to Advertising." The Marcel Dassault Award, Paris 1970.

_____. The Art of Using Science in Marketing, Harper and Row, New York 1974.

Ray, M.L., Sower, A.G., Strong, E.C.. "Frequency Effects Revisited." Journal of Advertising Research, Vol. II, February 1971, pp. 14-20.

_____. "Repetition in Media Models: A Laboratory Technique." Journal of Marketing Research, Vol. 8, February 1971, pp. 20-29.

_____. "Behavioral Measurements for Marketing Models: Estimating the Effects of Advertising Repetition for Media Planning." Management Science, Vol. 18, December 1971, pp. 73-89.

Rothman, L.J. "The use of Response Functions: A Discussion and an Alternative." The Thomson Medal and Awards for Media Research, London.1967, pp. 87-122.

"Round table 7: The Credibility Gap in Media Research." Media Decisions, April 1974.

Schaefer, W. "Scale Measures of Magazine Readership." Journal of Advertising Research, December 1965, Vol. 5.

Schonfeld, E.P. Prediction of Readership and Inquiry Response to Print Advertising. Unpublished Ph.D. Dissertation, Northwestern University, August 1975.

Sherill, Peter N. "The Probability of Advertising Exposure." Journal of Advertising Research, March 1966.

Schreiber, Robert J. "Letter to the Editor." Management Science, April 1968, B526-527.

_____. "Probability Assignments for the Simulation of Media Reach and Frequency." Journal of Advertising Research, Vol. 8, No. 2, June 1968.

_____. "The Metheringham Method for Media Mix: An Evaluation." Journal of Advertising Research, Vol. 9, No. 2, pp. 54-56, 1969.

_____. "Instability in Media Exposure Habits." Journal of Advertising Research, Vol. 14, No. 2, April 1974.

Simmons, W.R. and Associates Research, Inc. 1964 Study of Selective Markets, New York 1964.

Simmons, W.R. "New Look at Reach and Frequency." A.R.F. Proceedings of the 15th Annual Conference, October 1969.

_____. An Examination of Alternative Methods of Estimating Media Reach and Frequency. W.R. Simmons and Associates Research Inc., New York 1968.

Simmons, W.R. and Associates Research Inc. Selective Markets and the Media Reaching Them, New York 1969.

Simmons, W.R. "How to Measure Magazine Audiences." 19th Annual Conference of A.R.F., New York 1973.

Simulmatic Corporation. Simulmatic Media Mix. 1. General Description, 2. Technical Description, New York, October 1962.

Sissors, J.Z. Advertising Media Decision Making, Unpublished manuscript, Evanston 1966.

_____. "Matching Media with Markets." Journal of Advertising Research, Vol. II, October 1971, pp. 39-43.

Smith, H.A. "The Use of Panels in Readership Surveys." The Thomson Medals and Awards for Media Research, Paris 1966, pp. 9-27.

Standard Rate and Data Service. Dictionary of Terms Useful to Buyers of Advertising, 1959, Skokie, Illinois.

Starch, D. Reading Competitive Advertisements. Starch Advertisement Readership Service, No. 58, New York.

Starch, D. and Staff Inc. Starch Magazine Report, New York 1967.

_____. A Study of Questionning Procedures for Estimating Magazine Audiences, New York 1971.

Stasch, S.F. "Linear Programming and Space-Time Consideration in Media Selection." Journal of Advertising Research, Vol. 5, December 1965, pp. 40-46.

_____. "Linear Programming and Media Selection: A Comment." Journal of Marketing Research, May 1967, pp. 205-207.

Steyn, H.S. "On the Univariate Series $F_{(t)}=F(a_1, b_1, b_2, \ldots, b_k, c, t, t^2, \ldots, t^k)$ and its Applications to Probability Theory." Nederland Akademia Wetensch. Proceeding Series A, Vol. 58, pp. 588-95.

Stock, J.S. "A Comparison of Eight Audience Estimates." Journal of Advertising Research, September 1961.

Strong, E.K. "Size and Frequency in Advertising." Association of National Advertising Managers, Research Bulletin No. 4, April 1913.

Strong, E.C. "The Use of Field Experimental Observation in Estimating Advertising Recall." Journal of Marketing Research, Vol. II, November 1974, pp. 369-78.

Swanson, C.E. "Patterns of Nighttime Television Viewing." Proceedings of the American Marketing Association, June 1967, pp. 143-147.

_____. "The Frequency Structure of Television and Magazines." Journal of Advertising Research, Vol. 7, June 1967, pp. 8-14.

Summers, J.O. "Media Exposure Patterns of Consumer Innovators." Journal of Marketing, Vol. 36, January 1972, pp. 43-49.

Target GroupIndex. Introductory Volume, New York 1974.

Taylor, C.J. "Some Developments in the Theory and Application of Media Scheduling Methods." Operational Research Quarterly, Vol. 14, No. 3, September 1963, pp. 291-305.

Tigert, D.J. "A Psychographic Profile of Magazine Audiences: An Investigation of a Medium Climate." 1st Annual Conference, Association for Consumer Research, August 1969.

_____. "Are Television Audiences Really Different?" Proceedings of the American Marketing Association, April 1971, pp. 239-246.

The Value in Arbitron Radio, American Research Bureau, Inc., 1974.

Wells, W.D. "The Rise and Fall of Television Program Types." Journal of Advertising Research, Vol. 9, September 1964, pp. 21-27.

Wells, W.D., Banks, S., Tigert, D.J. "Order in the Data." Proceedings of the American Marketing Association, June 1967, pp. 263-267.

Wilson, C.L. "Homemaker Living Patterns and Market Place Behavior: A Psychometric Approach." Proceedings of the American Marketing Association, Spring 1966, pp. 305-331.

Wright, P.L. "On the Direct Monitoring of Cognitive Response to Advertising." Buyer/Consumer Information Processing, G.D. Hughes, H.L.Ray editors, University of North Carolina Press, 1974.

Young, L.F. "Advertising Media Models: Their Uses and Limitations." Computer Innovations in Marketing, American Marketing Association, 1970.

_____. "Estimating Radio Reach." Journal of Advertising Research, Vol. 12, October 1972, pp. 37-41.

"Young and Rubican, BBDO, Unleash Media Computerization." Advertising Age, October 1962.

Zangwill, W.I. "Media Selection by Decision Programming." Journal of Advertising Research, Vol. 5, September 1965, pp. 30-36.

Zielske, H.A. "The Remembering and Forgetting of Advertising." Journal of Marketing, Vol. 23, January 1959, pp. 240.